GMM Estimation of Short Dynamic Panel Data Models With Error Cross-Sectional Dependence

Sarafidis, Vasilis

University of Sydney

January 2009

GMM Estimation of Short Dynamic Panel Data Models With Error Cross-Sectional Dependence

Vasilis Sarafidis*

Preliminary and incomplete. January 2009

Abstract

This paper considers the issue of GMM estimation of a short dynamic panel data model when the errors are correlated across individuals. We focus particularly on the conditions required in the cross-sectional dimension of the error process for the dynamic panel GMM estimator to remain consistent. To this end, we demonstrate that cross-sectional independence (or uncorrelatedness) is not necessary − rather, it suffices that, if there is such correlation in the errors, this is weak. We define a stochastic scalar sequence to be cross-sectionally weakly correlated at any given point in time if the sequence of the covariances of the observations across individuals i and j at time t, given the conditioning set of all time-invariant characteristics of individuals i and j, converges absolutely as $N \rightarrow \infty$. Spatial dependence satisfies this condition but factor structure dependence does not. Consequently, the dynamic panel GMM estimator is consistent only in the first case. Under cross-sectionally weakly correlated errors, an additional, non-redundant, set of moment conditions becomes relevant for each i − specifically, instruments with respect to the individual(s) which unit i is correlated with. We demonstrate that these moment conditions remain valid when the errors are subject to both weak and strong correlations, in which situation the standard moment conditions with respect to individual i itself are invalidated − meaning that the dynamic panel GMM estimator is inconsistent. Simulated experiments show that the resulting method of moments estimators largely outperform the conventional ones in terms of both median bias and root median square error.

Key Words: dynamic panel data, spatial dependence, factor structure dependence, Generalised Method of Moments.
JEL Classification: C13; C31; C33.

1 Introduction

In developing the theory of GMM estimation of short dynamic panel data models, it is commonly assumed that the regression errors are independently distributed across

*Discipline of Econometrics and Business Statistics, University of Sydney, NSW 2006, Australia. Tel: +61-2-9036 9120; e-mail: v.sarafidis@econ.usyd.edu.au.

individuals (see e.g. Anderson and Hsiao, 1981, pg. 598, Arellano and Bond, 1991, pg. 278, Arellano, 1993, pg. 88, Ahn and Schmidt, 1995, pg. 7, Blundell and Bond, 1998, page 118, and others). This assumption is usually made for identification purposes rather than descriptive accuracy with the hope, presumably, that by conditioning on a sufficient number of explanatory variables, what is left over can be treated as a purely idiosyncratic disturbance that is uncorrelated across individuals. On the other hand, in empirical applications of GMM estimation this rather strong assumption is somewhat relaxed by allowing for common variations in the dependent variable at any given point in time using two-way error components disturbances (e.g. Arellano and Bond, 1991, pg. 288, Blundell and Bond, 1998, pg. 137, Bover and Watson, 2005, pg. 1975). In practice, however, a $\alpha_i + f_t + \varepsilon_{i,t}$ formulation is unlikely to be adequate to remove all correlated behaviour in the errors and this may result in misleading inferences and even inconsistent GMM estimators (Sarafidis and Robertson, 2008)[1].

Error cross-sectional dependence may arise for various reasons in practice; for example, it may be due to the presence of spatial correlations specified on the basis of economic and social distance (Conley, 1999) or relative location (Anselin, 1988), as well as due to the presence of unobserved components that give rise to a common factor specification in the disturbances with a fixed number of factors (e.g. Goldberger, 1972, and Jöreskog and Goldberger, 1975). Methods that account for a multi-factor error structure have been proposed by Robertson and Symons (2000), Coakley, Fuertes and Smith (2002), Phillips and Sul (2003), Moon and Perron (2004), Bai (2005), Pesaran (2006) and others. However, these methods are theoretically justified in panels where the number of time series observations (T) is large. To the best of our knowledge, no study exists that accounts for spatial correlations in a short dynamic panel data model.

The present paper deals specifically with the issue of GMM estimation of a short dynamic panel data model when the errors are not independent across individuals. A major focus lies on the conditions required in the cross-sectional dimension of the error process for the dynamic panel GMM estimator to remain consistent. To this end, we demonstrate that independence, or uncorrelatedness, is not necessary for GMM consistency or asymptotic efficiency − rather, it is sufficient that, if there is such correlation in the errors, this is weak. We define a stochastic scalar sequence to be cross-sectionally weakly correlated at any given point in time if the sequence of the covariances of the observations across individuals i and j at time t, given the conditioning set of all time-invariant characteristics of individuals i and j, converges absolutely as $N \to \infty$. Conversely, a sequence is strongly correlated if the sequence of the covariances does not converge absolutely. We show that the spatial approach to modelling error cross-sectional dependence, which typically assumes uniform boundedness of the row and column sums of the weighting matrix, satisfies weak correlation, although it is more restrictive in the sense that the latter does not require uniform boundedness. On the other hand, under factor structure dependence the errors are cross-sectionally strongly correlated and therefore the dynamic panel GMM estimator is not consistent. The two-way error components

[1] In an influential paper, Phillips and Sul (2007) analyse the impact of error cross section dependence on the dynamic Fixed Effects (FE) estimator.

model violates weak cross-sectional dependence too, albeit the problem can be dealt in this case via time-specific demeaning of the observations. However, careful analysis needs to be made in this case because the aforementioned transformation induces some dependency among the N individual equations and therefore the moment conditions are not valid anymore for finite N, a result that is usually ignored in the literature.

In addition, this paper shows that when the errors are cross-sectionally weakly correlated in the way defined above, then for each individual i there is an additional set of moment conditions that becomes relevant − in particular, instruments with respect to the individuals which unit i is spatially correlated with. We demonstrate that these extra moment conditions are not redundant in the sense that the asymptotic variance of the GMM estimator from the enlarged set of moment conditions is less than the GMM estimator that uses the smaller set of moment conditions, i.e. those instruments with respect to individual i. The spatial moment conditions can be particularly useful when the errors are subject to both weak and strong correlations because the standard moment conditions are invalidated in this case, meaning that the dynamic panel GMM estimator is inconsistent. The situation of both weak and strong correlations is also considered by Pesaran and Tosetti (2007) but for a model with no lags of the dependent variable on the right-hand side and T sufficiently large.

The structure of the paper is as follows. The following section specifies the panel regression model in a way that encompasses common unobserved factors and spatial dependence. Section 3 reviews the standard moment conditions used in GMM estimation under two-way error components disturbances. Section 4 addresses the issue of consistency for the dynamic panel GMM estimator when the independence assumption across individuals is relaxed. Section 5 shows that under cross-sectionally weakly correlated errors, additional non-redundant moment conditions become relevant for each individual i, which arise from the individual(s) which unit i is spatially correlated with. Section 6 demonstrates the validity of these extra moment conditions under both weakly and strongly correlated errors and the following section analyses the properties of the resulting GMM estimators, including when the problem of weak instruments applies. The performance of these estimators is investigated in Section 8 using simulated data. A final section concludes.

2 Model Specification

We focus on dynamic panel data models of the following first-order autoregressive form

$$\begin{aligned} y_{i,t} &= \lambda y_{i,t-1} + \upsilon_{i,t}, \quad i = 1, ..., N \text{ and } t = 2, ..., T \\ \upsilon_{i,t} &= \alpha_i + u_{i,t}, \;\; u_{i,t} = \sum_{m=1}^{M} \phi_{i,m} w_{i,m} f_{m,t} + \varepsilon_{i,t} = (\boldsymbol{\phi}_i \odot \mathbf{w}_i)' \mathbf{f}_t + \varepsilon_{i,t}, \end{aligned} \tag{1}$$

where $y_{i,t}$ is the observation of the dependent variable of the i^{th} cross-sectional unit at time t and λ is a fixed parameter to be estimated with $|\lambda| < 1$. $\upsilon_{i,t}$ is a composite error term that consists of α_i, an individual-specific time-invariant effect with zero mean and constant finite variance σ_α^2, and a weighted sum of error components, where $\mathbf{f}_t = (f_{1,t}, ..., f_{M,t})'$ denotes an $M \times 1$ vector of unobservables, $\boldsymbol{\phi}_i = (\phi_{i,1}, ..., \phi_{i,M})'$ is an $M \times 1$ vector of the corresponding coefficients, $\mathbf{w}_i = (w_{1,i}, ..., w_{M,i})'$ is an $M \times 1$ vector of deterministic bounded weights, $\odot$ denotes the Hadamard product and $\varepsilon_{i,t}$ is a purely idiosyncratic disturbance with zero mean and constant finite variance σ_ε^2.

We make the following assumptions

Assumption 1: $E(\alpha_i \varepsilon_{i,t}) = 0$ for all i, t, and $E(\alpha_i \alpha_j) = 0$ for all $i \neq j$.

Assumption 2: $E(\varepsilon_{i,t}\varepsilon_{i,s}) = 0$ for all i and $t \neq s$.

Assumption 3: $E(y_{i,1}u_{i,t}) = 0$, for $i = 1, ..., N$ and $t = 2, 3, ..., T$.

Assumption 4: $E(\mathbf{f}_t) = \mathbf{0}$, $E(\mathbf{f}_t\mathbf{f}_s') = \begin{cases} \boldsymbol{\sigma}_f^2 \mathbf{I}_M \text{ for } t = s, \text{ where } \boldsymbol{\sigma}_f^2 = \left(\sigma_{f_1}^2, ..., \sigma_{f_M}^2\right), \\ \mathbf{0} \text{ otherwise.} \end{cases}$

Assumption 5: $\phi_{i,m}$ is non-stochastic and bounded with $\lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^N \boldsymbol{\phi}_i\right] = \boldsymbol{\mu}_\phi$, $\lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^N \underline{\boldsymbol{\phi}}_i^o \underline{\boldsymbol{\phi}}_{i+k}^{o\prime}\right] = \begin{cases} \boldsymbol{\Sigma}_{\boldsymbol{\phi}} \text{ for } k = 0 \text{ and} \\ \mathbf{0} \text{ otherwise} \end{cases}$, where $\boldsymbol{\Sigma}_{\boldsymbol{\phi}}$ is an $M \times M$ positive semi-definite matrix and $\underline{\boldsymbol{\phi}}_i^o = \boldsymbol{\phi}_i - \boldsymbol{\mu}_{\boldsymbol{\phi}}$.

Assumption 6: $E(\alpha_i \mathbf{f}_t) = \mathbf{0}$, $E(\varepsilon_{it}\mathbf{f}_t) = \mathbf{0}$ for all i, t.

Assumptions 1-3 are standard in the GMM literature. Assumption 2 can be easily relaxed by allowing $\varepsilon_{i,t} \sim MA(k)$, where k is a small positive integer.[2] Assumption 3 ensures that sufficiently lagged values of $y_{i,t}$ will be uncorrelated with the first-difference of $\varepsilon_{i,t}$ and thus they will be available as instruments. Assumption 4 implies that $\mathbf{f}_t$ is serially and mutually uncorrelated. Assumption 5 ensures that the coefficients of $\mathbf{f}_t$ are bounded, as well as mutually and cross-sectionally uncorrelated. Assumption 6 is a random coefficients type of assumption and implies that the $\mathbf{f}_t$ are uncorrelated with α_i and $\varepsilon_{i,t}$ for all i and t.

Note that all the results discussed below extend in an obvious fashion to higher order autoregressive processes as well as to panel autoregressive distributed lag models. Define the $(T-1) \times M$ matrix $\mathbf{F} = [\mathbf{f}_2, \mathbf{f}_3, ...\mathbf{f}_T]'$, and the $N \times M$ matrices $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ...\boldsymbol{\phi}_N]'$, $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ...\mathbf{w}_N]'$. The initial model given in (1) can be written more compactly as

$$\mathbf{Y} = \lambda \mathbf{Y}_{-1} + \boldsymbol{\alpha} + \mathbf{u}, \;\; \mathbf{u} = (\boldsymbol{\Phi} \odot \mathbf{W})\,\mathbf{F}' + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{Y} = [\mathbf{Y}_1, ..., \mathbf{Y}_N]'$, a $N \times (T-1)$ matrix with $\mathbf{Y}_i = (y_{i,2}, y_{i,3}, ..., y_{i,T})'$, $\mathbf{Y}_{-1} = [\mathbf{Y}_{1,-1}, ..., \mathbf{Y}_{N,-1}]'$ a $N \times (T-1)$ matrix with $\mathbf{Y}_{i,-1} = (y_{i,1}, y_{i,2}, ..., y_{i,T-1})'$, $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_N]'$ with $\boldsymbol{\alpha}_i = \alpha_i \mathbf{i}_{T-1}$ and $\mathbf{i}_{T-1}$ being a $(T-1) \times 1$ column vector of ones and $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_N]'$ with $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,2}, \varepsilon_{i,3}, ..., \varepsilon_{i,T})'$.

[2] AR processes can be accomodated by adding further lags of the dependent variable on the right-hand side of the regression model.

The composite error term, $\mathbf{u}$, has a flexible structure in that it can characterise various forms of cross-sectional dependence, which includes dependence that is due to the presence of unobserved common factors, as well as spatial correlations in the error process, depending on the structure of $\mathbf{W}$. Specifically, the multi-factor structure arises from (2) by setting $\mathbf{W} = \mathbf{i}_M \mathbf{i}'_M$, where $\mathbf{i}_M$ is a $M \times 1$ column vector of ones. In this case we have $var\left(u_{i,t}\right) = \sum_{m=1}^{M} \phi_{i,m}^2 \sigma_{f_m}^2 + \sigma_\varepsilon^2$ and

$$cov\left(u_{i,t}, u_{i+k,s}\right) = \begin{cases} \sum_{m=1}^{M} \phi_{i,m}\phi_{i+k,m}\sigma_{f_m}^2 & \text{for } t = s, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

The Spatial Error Components (SEC) process[3] arises from (2) by imposing appropriate homogeneity restrictions on $\phi_{i,m}$, and setting $M = N$, $\sigma_{f_m}^2 = \sigma_f^2 \; \forall \; m$, and $\mathbf{W}$ equal to a sparse matrix populated primarily with zeros. For instance, in a circular[4] SEC(1) process $\mathbf{W}$ is given by

$$\mathbf{W} = \mathbf{W}^{SEC(1)} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & . & . & . & . & . & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & . & . & . & . & . & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & . & . & . & . & . & 0 \\ . & . & . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & . & . & . & 0 & 1 & 1 \\ 1 & 0 & 0 & . & . & . & . & . & . & 0 & 0 & 1 \end{bmatrix}, \quad (4)$$

and $\phi_{i,m} = \phi_{i(\mathrm{mod}\, N)+1, m(\mathrm{mod}\, N)+1} = \phi_\eta$ for $i = 1, ..., N$ and $m = i + \eta$, where $\eta = 0, 1$, while $n\,(\mathrm{mod}\, N)$ is the modulo operator, which is defined as the remainder after numerical division of n by N to obtain integer values. Thus, for $n = 1, ..., N-1$, $n\,(\mathrm{mod}\, N) + 1 = n + 1$, for $n = N$, $N\,(\mathrm{mod}\, N) + 1 = 1$, and for $n > N$, $n\,(\mathrm{mod}\, N) = n - N$. In this case we have $E\left(u_{i,t}\right) = 0$, $var\left(u_{i,t}\right) = \sigma_\varepsilon^2 + \left(\phi_0^2 + \phi_1^2\right)\sigma_f^2$ and

$$cov\left(u_{i,t}, u_{i+k,s}\right) = \begin{cases} \left(\phi_1 + \phi_2\right)\sigma_f^2 & \text{for } \kappa = i\,(\mathrm{mod}\, N) + 1 \text{ and } t = s, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

The Spatial Moving Average (SMA) process is a restricted case of the SEC form; it arises by setting $f_{m,t} = \varepsilon_{i,t}$ when $m = i$, for $m = 1, ..., N$. The structure of $\mathbf{W}$ depends again on the order of the spatial correlations. For example, the circular SMA(1) process sets

$$\mathbf{W} = \mathbf{W}^{SMA(1)} = \mathbf{W}^{SEC(1)} - \mathbf{I}_M. \quad (6)$$

In this case we have $E\left(u_{i,t}\right) = 0$, $var\left(u_{i,t}\right) = \sigma_\varepsilon^2\left(1 + \phi_1^2\right)$ and

$$cov\left(u_{i,t}, u_{i+k,s}\right) = \begin{cases} \phi_1\sigma_\varepsilon^2 & \text{for } \kappa = i\,(\mathrm{mod}\, N) + 1 \text{ and } t = s, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

[3] See Kelejian and Robinson (1995).
[4] See e.g. Baltagi, Bresson and Pirotte (2007).

SMA processes of higher order can be accomodated in a straightforward way. Assuming invertibility, the Spatial Autoregressive (SAR) form can be obtained using an infinite SMA representation[5].

Hence, spatial dependence can be viewed as a special form of factor structure dependence with, essentially, appropriate zero restrictions on $\mathbf{W}$ and homogeneity restrictions on $\mathbf{\Phi}$. Therefore, one may think of the unobserved components, $\mathbf{F}$, in (2) as shocks, the impact of which is either 'global' (factors) or 'local' (spatially correlated components), depending on the structure of $\mathbf{W}$. It follows that mixture cases can be accomodated in a straightforward way; for example, a K-factor process along with a spatial first-order moving average process arises from (2) by setting $M = N + K$, $\phi_{i,m} = \phi_{i(\text{mod}\, N)+1,m(\text{mod}\, M)+1} = \phi$ for $i = 1, ..., N$ and $m = i+1$, $f_{m,t} = \varepsilon_{i,t}$ when $m = i$, for $m = 1, ..., N$, and finally

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & . & . & . & . & . & 0 & 1 & \cdots & 1 \\ 0 & 0 & 1 & 0 & . & . & . & . & 0 & 1 & \cdots & 1 \\ . & . & . & . & . & . & . & . & . & 1 & \cdots & 1 \\ . & . & . & . & . & . & . & . & . & 1 & \cdots & 1 \\ 0 & . & . & . & . & 0 & 0 & 0 & 1 & 1 & \cdots & 1 \\ 1 & . & . & . & . & . & 0 & 0 & 0 & 1 & \cdots & 1 \end{bmatrix}. \tag{8}$$

We consider estimation of mixture models in Section 6.

3 Moment Conditions in Standard GMM Estimation

Typical GMM estimation of linear dynamic panel data models of the form given in (1) imposes $\phi_{i,m} = 0$ for all i and m, such that any form of dependence in the error process across individuals, whether this is spatial or subject to a factor structure, is ruled out[6]. Consequently, applying first-differences in (1) yields

$$\Delta y_{i,t} = \lambda \Delta y_{i,t-1} + \Delta \upsilon_{i,t}, \quad i = 1, ..., N \text{ and } t = 3, ..., T. \tag{9}$$

Using Assumptions 1-3 the following $\zeta = (T-1)(T-2)/2$ moment conditions become available

$$E(y_{i,t-s}\Delta \upsilon_{i,t}) = 0; \text{ for } t = 3, ..., T \text{ and } 2 \leq s \leq t-1. \tag{10}$$

On the other hand, in empirical applications it is common practice to generalise the error structure by allowing for common variations in the dependent variable using a two-way error components formulation[7]:

$$\upsilon_{i,t} = \alpha_i + f_t + \varepsilon_{i,t}. \tag{11}$$

[5] In this case, $\mathbf{W}$ is not sparse, however its elements will decline with a distance measure that increases sufficiently rapidly as the sample increases. For instance, Stetzer (1982) models the distance decay by a negative exponential function, $w_{i,m} = \exp(-\theta d_{i,m})$, $0 < \delta < \infty$, with $d_{i,m}$ denoting the distance between individuals i and m.

[6] See Section 1 for related references.

[7] Viz. footnote 5.

The above formulation can be viewed as a degenerate factor structure in which $\phi_i = \phi = 1$. In this case, the moment conditions given in (10) are not valid anymore because for fixed T we have[8]

$$\begin{aligned} & E\left(y_{i,t-s}\Delta\upsilon_{i,t} | \left\{f_n\right\}_{-\infty}^{t}\right) = \\ = \; & E\left[\left(\frac{\alpha_i}{1-\lambda} + \sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{i,t-s-\tau} + \sum_{\tau=0}^{\infty}\lambda^{\tau}f_{t-s-\tau}\right)\left(\Delta f_t + \Delta\varepsilon_{i,t}\right) | \left\{f_n\right\}_{-\infty}^{t}\right] \\ = \; & \sum_{\tau=0}^{\infty}\lambda^{\tau}f_{t-s-\tau}\Delta f_t \neq 0. \end{aligned} \tag{12}$$

Notice that in our large N asymptotics it is the conditional expectation that is relevant since we are never taking large T averages. Transforming the observations in terms of deviations from time-specific averages eliminates this problem by removing the common time effect from the regression error

$$\underline{\upsilon}_{it} = \upsilon_{it} - \overline{\upsilon}_t = (\alpha_i - \overline{\alpha}) + (f_t - f_t) + (\varepsilon_{it} - \overline{\varepsilon}_t) = \underline{\alpha}_i + \underline{\varepsilon}_{it}. \tag{13}$$

Of course, this transformation induces some dependency across individuals and therefore the moment conditions on the transformed observations are valid only for large N, which is what we require. In particular, defining $\underline{y}_{i,t-s} = y_{i,t-s} - \overline{y}_{t-s}$ with $\overline{y}_{t-s} = \frac{1}{N}\sum_{i}^{N} y_{it-s}$, $\underline{y}^{o}_{i,t-s} = y_{i,t-s} - \widetilde{\mu}_{y_{t-s}}$, where $\widetilde{\mu}_{y_{t-s}} = E\left(y_{i,t-s} | \left\{f_n\right\}_{-\infty}^{t-s}\right)$, $\Delta\underline{\upsilon}_{it} = \Delta\left[\upsilon_{i,t} - \overline{\upsilon}_t\right]$ and $\Delta\underline{\upsilon}^{o}_{it} = \Delta\left[\upsilon_{i,t} - \widetilde{\mu}_{\upsilon_t}\right]$, where $\widetilde{\mu}_{\upsilon_t} = E\left(\upsilon_{i,t} | \left\{f_n\right\}_{-\infty}^{t}\right)$, we have

$$\begin{aligned} & \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\left[\underline{y}_{i,t-s}\Delta\underline{\upsilon}_{i,t}\right] \\ = \; & \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\left[\left(y_{i,t-s} - \widetilde{\mu}_{y_{t-s}}\right) - \left(\overline{y}_{t-s} - \widetilde{\mu}_{y_{t-s}}\right)\right]\Delta\left[\left(\upsilon_{i,t} - \widetilde{\mu}_{\upsilon_t}\right) - \left(\overline{\upsilon}_t - \widetilde{\mu}_{\upsilon_t}\right)\right] \\ = \; & \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\underline{y}^{o}_{i,t-s}\Delta\underline{\upsilon}^{o}_{i,t}\right] + o_p(1), \end{aligned} \tag{14}$$

since $\Delta\left(\overline{\upsilon}_t - \widetilde{\mu}_{\upsilon_t}\right) = O_p\left(N^{-1/2}\right)$, $\overline{y}_{t-s} - \widetilde{\mu}_{y_{t-s}} = O_p\left(N^{-1/2}\right)$, $N^{-1/2}\sum_{i=1}^{N}\underline{y}^{o}_{i,t-s} = O_p(1)$ and $N^{-1/2}\sum_{i=1}^{N}\Delta\underline{\upsilon}^{o}_{it} = O_p(1)$. Hence, given that $\underline{y}^{o}_{i,t-s}\Delta\underline{\upsilon}^{o}_{i,t}$ are independent across i, a suitable CLT (Central Limit Theorem) ensures that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\underline{y}^{o}_{i,t-s}\Delta\underline{\upsilon}^{o}_{i,t}\right] \xrightarrow{d} N\left(0, var\left(\frac{1}{\sqrt{N}}\underline{y}^{o}_{i,t-s}\Delta\underline{\upsilon}^{o}_{i,t}\right)\right). \tag{15}$$

[8] See also Sarafidis and Robertson (2008).

Now define

$$\underline{\mathbf{Z}}_i = \begin{bmatrix} \underline{y}_{i,1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \underline{y}_{i,1} & \underline{y}_{i,2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \underline{y}_{i,1} & \underline{y}_{i,2} & \cdots & \underline{y}_{i,T-2} \end{bmatrix};\ \Delta\underline{\boldsymbol{\upsilon}}_i = \begin{bmatrix} \Delta\underline{\upsilon}_{i,3} \\ \Delta\underline{\upsilon}_{i,4} \\ \vdots \\ \Delta\underline{\upsilon}_{i,T} \end{bmatrix}, \tag{16}$$

and $\underline{\mathbf{Z}} = (\underline{\mathbf{Z}}_1', \underline{\mathbf{Z}}_2', ..., \underline{\mathbf{Z}}_N')'$. The cross-sectionally demeaned and first-differenced GMM estimator equals

$$\widehat{\lambda}_{DIF\ GMM} = \left[\frac{1}{N}\Delta\underline{\mathbf{y}}_{-1}'\underline{\mathbf{Z}}\widehat{\mathbf{A}}_N\underline{\mathbf{Z}}'\Delta\underline{\mathbf{y}}_{-1}\right]^{-1}\left[\frac{1}{N}\Delta\underline{\mathbf{y}}_{-1}'\underline{\mathbf{Z}}\widehat{\mathbf{A}}_N\underline{\mathbf{Z}}'\Delta\underline{\mathbf{y}}\right] \tag{17}$$

with $\Delta\underline{\mathbf{y}} = \left(\Delta\underline{\mathbf{y}}_1', ..., \Delta\underline{\mathbf{y}}_N'\right)'$, $\Delta\underline{\mathbf{y}}_i = \left(\Delta\underline{y}_{i,3}, ..., \Delta\underline{y}_{i,T}\right)'$, $\Delta\underline{\mathbf{y}}_{,-1} = \left(\Delta\underline{\mathbf{y}}_{1,-1}', ..., \Delta\underline{\mathbf{y}}_{N,-1}'\right)'$ and $\Delta\underline{\mathbf{y}}_{i,-1} = \left(\Delta\underline{y}_{i,2}, ..., \Delta\underline{y}_{i,T-1}\right)'$. $\widehat{\mathbf{A}}_N$ is some weighting matrix that satisfies

$$\widehat{\mathbf{A}}_N - \mathbf{A}_N \xrightarrow{p} 0, \tag{18}$$

where $\mathbf{A}_N$ is a non-stochastic sequence of positive definite matrices. Alternative choices of $\widehat{\mathbf{A}}_N$ lead to different GMM estimators, which are all consistent but they differ in terms of efficiency. The asymptotically efficient DIF GMM estimator sets $\widehat{\mathbf{A}}_N$ equal to the inverse of the covariance matrix of the moment conditions[9] – that is, $\widehat{\mathbf{A}}_N^{-1} = est.asy.var\left(\sqrt{N}\frac{1}{N}\underline{\mathbf{Z}}'\boldsymbol{\Delta}\underline{\boldsymbol{\upsilon}}\right)$, assuming that this matrix exists and is finite positive definite. When ε_{it} is homoskedastic $\widehat{\mathbf{A}}_N^{-1}$ can be approximated by

$$\widehat{\mathbf{A}}_N^{-1} = \frac{1}{N}\underline{\mathbf{Z}}'\mathbf{H}\underline{\mathbf{Z}}, \tag{19}$$

where

$$\mathbf{H} = \mathbf{I}_N \otimes \mathbf{H}_i \tag{20}$$

and

$$\mathbf{H}_i = \begin{bmatrix} 2 & -1 & \cdots & & 0 \\ -1 & 2 & \cdots & & 0 \\ \vdots & \vdots & & & \vdots \\ . & . & & & -1 \\ 0 & 0 & \cdots & -1 & 2 \end{bmatrix}. \tag{21}$$

Therefore, it is clear that the weighting matrix $\widehat{\mathbf{A}}_N^{-1}$ involves the Kronecker product between two distinct matrices, the former of which reflects cross-sectional dependence in the error structure (which, for large N, is zero in the present case and hence the use of the identity matrix) while the latter reflects time series dependence in the error structure,

[9] See Hansen (1982).

and in particular first-order serial correlation, which is induced by first-differencing the observations. Note that since the individual equations are independent across i for large N, $\widehat{\mathbf{A}}_N^{-1}$ can also be written as $\widehat{\mathbf{A}}_N^{-1} = \frac{1}{N}\sum_i \underline{\mathbf{Z}}_i'\mathbf{H}_i\underline{\mathbf{Z}}_i$. Therefore, an equivalent expression for (17) is given by

$$\widehat{\lambda}_{DIF\ GMM} = \left[\frac{1}{N}\left(\sum_{i=1}^{N}\Delta\underline{\mathbf{y}}_{i,-1}'\underline{\mathbf{Z}}_i\right)\left(\frac{1}{N}\sum_i \underline{\mathbf{Z}}_i'\mathbf{H}_i\underline{\mathbf{Z}}_i\right)^{-1}\left(\sum_{i=1}^{N}\underline{\mathbf{Z}}_i'\Delta\underline{\mathbf{y}}_{i,-1}\right)\right]^{-1} \left[\frac{1}{N}\left(\sum_{i=1}^{N}\Delta\underline{\mathbf{y}}_{i,-1}'\underline{\mathbf{Z}}_i\right)\left(\frac{1}{N}\sum_i \underline{\mathbf{Z}}_i'\mathbf{H}_i\underline{\mathbf{Z}}_i\right)^{-1}\left(\sum_{i=1}^{N}\underline{\mathbf{Z}}_i'\Delta\underline{\mathbf{y}}_{i}\right)\right]. \quad (22)$$

When the individual observations are not independent across i, (22) is not equivalent to (17).

The standard first-differenced GMM (DIF GMM) estimator may have poor finite sample properties in terms of bias and precision when $\lambda \to 1$ or $\sigma_\alpha^2/\sigma_\varepsilon^2 \to \infty$. As a result, Blundell and Bond (1998) developed an approach outlined in Arellano and Bover (1995), which combines the equations in first-differences with the equations in levels, using $\Delta y_{i,t-1}$ as an instrument for the lagged dependent variable, $y_{i,t-1}$

$$E\left(\Delta y_{i,t-1}\upsilon_{i,t}\right) = 0; \text{ for } t = 3, 4, ..., T. \quad (23)$$

This approach gives rise to a system GMM (SYS GMM) estimator, which is valid provided that the deviations of the initial observations from the long-run convergent values are uncorrelated with the individual-specific, time-invariant effects − that is,

$$E\left[\alpha_i\left(y_{i,1} - \frac{\alpha_i}{1-\lambda}\right)\right] = 0. \quad (24)$$

If common time effects are included in the error process, what is required is that

$$\text{plim}_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}\left[\underline{\alpha}_i\left(\underline{y}_{i,1} - \frac{\alpha_i}{1-\lambda}\right)\right]\right] = 0. \quad (25)$$

Thus, defining

$$\underline{\mathbf{Z}}_i^{sys} = \begin{bmatrix} \underline{\mathbf{Z}}_i & 0 & \cdots & 0 \\ 0 & \Delta\underline{y}_{i,2} & & 0 \\ . & . & \ddots & . \\ 0 & 0 & \cdots & \Delta\underline{y}_{i,T} \end{bmatrix}; \ \underline{\boldsymbol{\upsilon}}_i^{sys} = \begin{bmatrix} \Delta\underline{\boldsymbol{\upsilon}}_i \\ \underline{\boldsymbol{\upsilon}}_i \end{bmatrix},$$

and $\underline{\mathbf{Z}}^{sys} = \left(\underline{Z}_1^{sys\prime}, \underline{Z}_2^{sys\prime}, ..., \underline{Z}_N^{sys\prime}\right)'$, the SYS GMM estimator is given by

$$\widehat{\lambda}_{SYS\ GMM} = \left[\frac{1}{N}\underline{\mathbf{y}}_{-1}^{sys\prime}\underline{\mathbf{Z}}^{sys}\widehat{\mathbf{A}}_{N,sys}\underline{\mathbf{Z}}^{sys\prime}\underline{\mathbf{y}}_{-1}^{sys}\right]^{-1}\left[\frac{1}{N}\underline{\mathbf{y}}_{-1}^{sys\prime}\underline{\mathbf{Z}}^{sys}\widehat{\mathbf{A}}_{N,sys}\underline{\mathbf{Z}}^{sys\prime}\underline{\mathbf{y}}^{sys}\right] \quad (26)$$

where $\underline{\mathbf{y}}^{sys} = \left(\underline{\mathbf{y}}_1^{sys\prime}, \underline{\mathbf{y}}_2^{sys\prime}, ..., \underline{\mathbf{y}}_N^{sys\prime}\right)'$, $\underline{\mathbf{y}}_i^{sys} = \left(\Delta \underline{y}_{i,3}, ..., \Delta \underline{y}_{i,T}, \underline{y}_{i,3}, ..., \underline{y}_{i,T}\right)'$, $\underline{\mathbf{y}}_{-1}^{sys} = \left(\underline{\mathbf{y}}_{1,-1}^{sys\prime}, \underline{\mathbf{y}}_{2,-1}^{sys\prime}, ..., \underline{\mathbf{y}}_{N,-1}^{sys\prime}\right)'$, and $\underline{\mathbf{y}}_{i,-1}^{sys} = \left(\Delta \underline{y}_{i,2}, ..., \Delta \underline{y}_{i,T-1}, \underline{y}_{i,2}, ..., \underline{y}_{i,T-1}\right)'$.

With homoskedastic errors the optimal choice of $\widehat{\mathbf{A}}_{N,sys}$ is given by[10]

$$\widehat{\mathbf{A}}_{N,sys} = \left[\frac{1}{N}\underline{\mathbf{Z}}^{sys\prime}\mathbf{H}^{sys}\underline{\mathbf{Z}}^{\mathbf{sys}}\right]^{-1}, \tag{27}$$

where $\mathbf{H}^{sys}$ equals

$$\mathbf{H}^{sys} = \begin{bmatrix} \mathbf{H} & \mathbf{D} \\ \mathbf{D}' & \mathbf{I}_{N(T-2)} \end{bmatrix} \tag{28}$$

with $\mathbf{I}_{N(T-2)} = \mathbf{I}_N \otimes \mathbf{I}_{T-2}$ and $\mathbf{C} = \mathbf{I}_N \otimes \mathbf{D}_1$, where $\mathbf{D}_1$ takes the value of 1 on the main diagonal, -1 on the first lower off-diagonal and zero otherwise.

The next section addresses the issue of consistency for the dynamic panel GMM estimator when the restriction $\phi_{i,m} = 0 \ \forall \ i, \ m$ is relaxed.

4 The Consistency of the Dynamic Panel GMM Estimator Under Error Cross-sectional Dependence

Given Assumptions 1-6 and model (2), the structure of $\mathbf{W}$ will be critical upon the asymptotic properties of the GMM estimator. To begin with, we firstly define the concept of a cross-sectionally weakly correlated process. Let $\left\{\upsilon_i^t, \ i \geq 1\right\}$ be the scalar sequence $\upsilon_{1,t}, \upsilon_{2,t}, \upsilon_{3,t}, ...$ There are $(T-1)$ such scalar sequences, for $t = 2, ..., T$.

Definition 1 *The double-indexed sequence $\{\upsilon_{it}, \ i \geq 1, \ t \geq 1\}$ is said to be cross-sectionally weakly correlated if, for each i and $j > 0$, $\{\sigma_{ij,t}, \ j \neq i\}$ converges absolutely as $N \to \infty$, that is,*

$$\sum_{j \neq i} |\sigma_{ij,t}| < \infty, \text{ for all } t, \tag{29}$$

where $\sigma_{ij,t} = Cov(\upsilon_{it}, \upsilon_{jt}|\Upsilon_{ij})$, and Υ_{ij} denotes the conditioning set of all time-invariant characteristics of individuals i and j. Similarly, the same sequence is said to be cross-sectionally strongly correlated if it does not converge absolutely, i.e. (29) *is violated.*

Theorem 2 *Let $\left\{\upsilon_i^t, \ i \geq 1\right\}$ be the scalar sequence $\upsilon_{1,t}, \upsilon_{2,t}, \upsilon_{3,t}, ...,$ and satisfies Assumptions 1-6, where $\upsilon_{i,t} = \alpha_i + \sum_{m=1}^{M} \phi_{i,m} w_{i,m} f_{m,t} + \varepsilon_{i,t}$. Suppose that (i) $w_{i,m} = O(1)$ $\forall \ i, \ m$, (ii) $R_i/N \to 1$ for $i = 1, ..., N$, where R_i is the number of elements in the i^{th} row of $\mathbf{W}$ that are either zero or at most of order $N^{-1/2-\eta}$ for some $\eta \geq 0$, and (iii) $C_m/N \to 1$, where C_m is the number of elements in the m^{th} column of $\mathbf{W}$ that are either zero, or of smaller order than $N^{-1/2}$. Then $\left\{\upsilon_i^t\right\}$ is cross-sectionally weakly correlated.*

[10] This is for $\sigma_\alpha^2 = 0$; see Windmeijer (2000) and Kiviet (2007).

Proof. See Appendix A. ■

Note that Theorem 2 is more general than a uniform boundedness condition for the row and column sums of $\mathbf{W}$ (typically employed in spatial models), which is stated as follows[11]

$$\sum_{i=1}^{N} |w_{i,m}| \leq B_w < \infty \ \forall \ m \text{ and } \sum_{m=1}^{M} |w_{i,m}| \leq B_w < \infty \ \forall \ i, \ M = N. \tag{30}$$

This is because uniform boundedness satisfies conditions (i)-(iii) of Theorem 2 but not vice versa. In particular, under Theorem 2 it is possible that $||\mathbf{W}||_{\infty} = \max_i \sum_{m=1}^{M} |w_{i,m}| = O\left(N^{1/2}\right)$ and $||\mathbf{W}||_1 = \max_m \sum_{i=1}^{N} |w_{i,m}| = o\left(N^{1/2}\right)$ – hence, the row/column sums of $\mathbf{W}$ will not necessarily be bounded, even if the error process is weakly correlated. This situation may arise in a number of empirical applications, where the effect of some factors depends on the size of N. For example, in the context of price competition among firms, the impact of a common shock may depend on the number of firms in the industry; thus, while for fixed N this impact may be non-zero, as N grows large, higher competition ensures that firms absorb the shock through a decrease in profits, as opposed to setting higher prices.

In conclusion, a spatially correlated process is cross-sectionally weakly correlated. On the other hand, the factor structure sets $\mathbf{W} = \mathbf{i}_M \mathbf{i}'_M$, where $\mathbf{i}_M$ is a $M \times 1$ column vector of ones, and so it violates conditions (ii) and (iii). Hence it provides an example of a process that is not weakly correlated. As a matter of fact, under a factor structure there are M unobserved variables, $\mathbf{f}_t$, which are common for all i and therefore their effect does not diminish no matter how far in the sequence two random variables, $\upsilon_{i,t}$ and $\upsilon_{j,t}$, are. As a result, the factor structure dependence is an example of a cross-sectionally strongly correlated process. The two-way error components model is a degenerate case of the single-factor structure because it sets $\phi_i = \phi$ for all i. Therefore, it provides another example of a strongly correlated process, albeit the correlation can be removed in this case for large N by transforming the data in terms of deviations from time-specific averages.

Notice that conditions (i)-(iii) of Theorem 2 do not imply that the $\left\{\upsilon_i^t, \, i \geq 1\right\}$ sequence is spatially ergodic because the row sums of $\mathbf{W}$ need not necessarily be the same, in which case the elements of the sequence are not identically distributed. Furthermore, the conditions do not require that the sequence is a mixing process either, in the sense that the elements of the sequence can be asymptotically uncorrelated but not asymptotically independent[12].

Remark 3 *Pesaran and Tosetti (2007) define the scalar sequence* $\left\{z_i^t, \, i \geq 1\right\}$ *to be weakly dependent at any given* t *if its (weighted) average converges to its expectation in quadratic mean. Specifically, let* w_i^{t-1} *denote a weight that satisfies* $\sum_i \left(w_i^{t-1}\right)^2 =$

[11] See e.g. Kapoor, Kelejian and Prucha (2007, pg. 106) and Lee (2007, pg. 491).

[12] Of course, this requires a strengthening of the moment restrictions – namely, $E\,|u_i|^2 < B_u < \infty$, as opposed to – say – $E\,|u_i|^c < B_u < \infty$ for $c > 1$).

$O\left(N^{-1/2}\right)$ and $w_i^{t-1}\left[\sum_i \left(w_i^{t-1}\right)^2\right]^{-1} = O\left(N^{-1/2}\right)$ for any $i \leq N$, and let I_{t-1} be the information set at time $t-1$ containing at least $\mathbf{z}^{t-1}$, $\mathbf{z}^{t-2}$,... and $\mathbf{w}^{t-1}$, $\mathbf{w}^{t-2}$,..., where $\mathbf{z}^{t-1} = \left(z_1^{t-1}, ..., z_N^{t-1}\right)'$ and $\mathbf{w}^{t-1} = \left(w_1^{t-1}, ..., w_N^{t-1}\right)'$. Then the sequence $\left\{z_i^t,\, i \geq 1\right\}$ is weakly dependent if

$$\lim_{N\to\infty} var\left(\sum_i w_i^{t-1} z_i^t \middle| I_{t-1}\right) = 0.$$

Under this definition, the following single-factor structure process

$$\begin{aligned} & u_{i,t} = \phi_i f_t + \varepsilon_{i,t}, \\ & \text{where } \phi_i \text{ is non-stochastic and bounded, and} \\ & E\left(f_t\right) = E\left(\varepsilon_{i,t}\right) = 0,\ E\left(f_t^2\right) = \sigma_f^2,\ E\left(\varepsilon_{i,t}^2\right) = \sigma_f^2,\ E\left(f_t \varepsilon_{i,t}\right) = 0, \end{aligned} \tag{31}$$

is weakly correlated so long as $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \phi_i = 0$[13]. This is not the case, however, using Definition 1 since it is straightforward to show that $\sum_{j\neq i}\left|\sigma_{ij,t}\right|$ is not bounded in this case. Intuitively, since all individuals are subject to the same shock the sum of the absolute conditional covariances between individual disturbances grows with N.

The following theorem provides a law of large numbers for weakly correlated sequences.

Theorem 4 *Let $\left\{\upsilon_i^t,\, i \geq 1\right\} = \upsilon_{1,t}, \upsilon_{2,t}, \upsilon_{3,t}, ...$ be a scalar sequence with cross-sectionally weakly correlated elements, such that $E\left(\upsilon_{i,t}\right) = 0$ and $E\left(\upsilon_{i,t}^2\right) < B_\upsilon < \infty$. Then*

$$\frac{1}{N}\sum_{i=1}^N \upsilon_{i,t} - E\left(\upsilon_{i,t}\right) \xrightarrow{p} 0 \tag{32}$$

Proof. It follows directly from Stout (1974, Corollary 2.4.1) and the Kronecker lemma. ■

Theorem 2 shows that so long as conditions (i)-(iii) are satisfied, $\upsilon_{i,t}$ is weakly correlated, or asymptotically uncorrelated across i. In turn, according to Theorem 4, the latter implies that the first sample centered moment of $\upsilon_{i,t}$ will converge in probability to zero. The following corollary provides the additional condition necessary to validate the moment conditions given in (10) and (23) under weakly correlated errors.

Corollary 5 *Let $\left\{\upsilon_i^t,\, i \geq 1\right\}$ and $\left\{u_i^t,\, i \geq 1\right\}$ be two scalar sequences, $\upsilon_{1,t}, \upsilon_{2,t}, \upsilon_{3,t}, ...$ and $u_{1,t}, u_{2,t}, u_{3,t}, ...$ respectively, which are cross-sectionally weakly correlated. The product of these sequences will also be cross-sectionally weakly correlated.*

[13] See Pesaran and Tosetti (2007), Theorem 16, page 15.

Direct application of Corollary 5 implies that

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(y_{i,t-s}\Delta\upsilon_{i,t}) = \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{\alpha_i}{1-\lambda}+\sum_{\tau=0}^{\infty}\lambda^{\tau}u_{i,t-s-\tau}\right)\Delta u_{i,t}\right]$$

$$= \sum_{\tau=0}^{\infty}\lambda^{\tau}\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}u_{i,t-s-\tau}\Delta u_{i,t} = 0 \text{ for } t=3,...,T \text{ and } s=2,...,T-1, \quad (33)$$

and

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(\Delta y_{i,t-1}\upsilon_{i,t}) = \sum_{\tau=0}^{\infty}\lambda^{\tau}\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Delta u_{i,t-1-\tau}\upsilon_{it} = 0.\text{for } t=3,4,...,T, \quad (34)$$

under weakly correlated errors. Therefore DIF GMM and SYS GMM are consistent.

Remark 6 *Observe that when a weakly correlated process is defined as in Remark* 3*, the sample average over* i *of the product between* $\{\upsilon_{i,t-s-\tau},\ i\geq 1\}$ *and* $\{\Delta\upsilon_{i,t},\ i\geq 1\}$ *does not necessarily converge to its expectation; for instance, for the single-factor process given in* (31) *and assuming that* $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\phi_i=0$*, we have* $\frac{1}{N}\sum_{i=1}^{N}u_{i,t-s}-E(u_{i,t-s})\to 0$ *and* $\frac{1}{N}\sum_{i=1}^{N}\Delta u_{i,t}-E(\Delta u_{i,t})\to 0$*. However, the sample average* $\frac{1}{N}\sum_{i=1}^{N}u_{i,t-s}\Delta u_{i,t}$ *converges to* $\mu_{\phi}^2 f_{t-s}\Delta f_t$*, where* $\mu_{\phi}^2 = plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\phi_i^2$*, despite the fact that* $E(u_{i,t-s}\Delta u_{i,t})=0$ *for* $s=2,...,t-1$*.*

Since asymptotic uncorrelatedness encompasses spatial dependence, it follows that DIF GMM and SYS GMM are consistent under spatially correlated errors. On the other hand, under factor structure dependence the correlation between $\upsilon_{i,t}$ and $\upsilon_{ij,t}$ persists. Therefore the law of large numbers provided above breaks down and $\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}y_{i,t-s}\upsilon_{i,t}\neq 0$, despite the fact that $E(y_{i,t-s}\upsilon_{i,t})=0$[14].

In conclusion, it has been shown that the dynamic panel GMM estimator does not require cross-sectionally independent errors for consistency − rather, it suffices that, if there is such dependence, this is weak (in the way defined above) at any given point in time. Theorem 2 shows that this holds true under conditions (i)-(iii), which are more general than uniform boundedness of the row and column sums of $\mathbf{W}$. The factor structure in the error process violates these conditions and therefore the standard panel GMM estimator is not consistent in this case.

5 Moment Conditions With Spatial Dependence

Suppose that the errors are spatially correlated. It turns out that not only DIF GMM and SYS GMM are consistent, but also, there is an additional set of moment conditions

[14] See also Sarafidis and Robertson (2008).

that becomes relevant in this case. These moment conditions are non-redundant in the sense that they increase the asymptotic efficiency of the estimators.[15]

In particular, consider the basic model in (1) and let $M = N$, which yields

$$y_{i,t} = \lambda y_{i,t-1} + \upsilon_{i,t},\, \upsilon_{i,t} = \alpha_i + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)' \mathbf{f}_t + \varepsilon_{i,t},\;\; i = 1, ..., N, \text{ and } t = 2, ..., T, \quad (35)$$

where $\boldsymbol{\phi}_i = \left(\phi_{i,1}, ..., \phi_{i,M}\right)'$ is the i^{th} row of $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_N]'$ and $\mathbf{\Phi}$ is an $N \times M$ $[= N]$ matrix with $\phi_{i,m} = \phi_{i(\text{mod}\, N)+1, m(\text{mod}\, N)+1} = \phi_\eta$ for $i = 1, ..., N$ and $m = i + \eta$, where $\eta = 0, 1, ..., N-1$. Also, $\mathbf{w}_i = (w_{i,1}, ..., w_{i,M})'$ is the i^{th} row of $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N]'$, where $\mathbf{W}$ is a sparse weighting matrix populated primarily with zeros. Let $\widetilde{\mathbf{y}}_{i,t}$ be a vector that contains the non-zero elements of $[(y_{1,t}, y_{2,t}, ..., y_{N,t}) \odot \mathbf{w}_i']'$ but excludes y_{it} itself, i.e. a vector that contains those elements that are spatially correlated with y_{it}, depending on the structure of $\mathbf{w}_i$. Under Assumptions 1-6, the autoregressive model (35) implies the following additional $\zeta\,(T-1)\,(T-2)\,/2$ moment conditions in the first-differenced equations

$$E\left(\widetilde{\mathbf{y}}_{i,t-s} \Delta u_{i,t}\right) = 0;\;\text{ for } t = 3, ..T \text{ and } 2 \leq s \leq t-1, \quad (36)$$

and the following $\zeta\,(T-2)$ moment conditions in the equations in levels

$$E\left(\Delta \widetilde{\mathbf{y}}_{i,t-1} u_{i,t}\right) = 0;\;\text{ for } t = 3, ..., T, \quad (37)$$

where ζ denotes the number of non-zero elements in $\mathbf{w}_i$. Notice that for the moment conditions in (37) we have not imposed any restrictions on the initial conditions process generating y_{i1} – other than the usual one provided by Assumption 3 of course. On the contrary, when $\Delta y_{i,t-1}$ is used as an instrument for $y_{i,t-1}$, (24) is required.

Theorem 7 *The spatial moment conditions* (36) *and* (37) *are non-redundant in the sense that the GMM estimators that use the enlarged set of moment conditions,* (10) *together with* (36) *and* (23) *together with* (37)*, have smaller asymptotic variance than the GMM estimators that use the smaller set,* (10) *and* (23)*.*

Proof. See Appendix B. ■

Thus, for example, for an SMA(1) process we have

$$\begin{aligned}
& cov\left[\widetilde{\mathbf{y}}_{i,t-s}, (\Delta y_{i,t-1} - \widehat{\pi}_1 y_{i,t-s})\right] = \\
= \;& \frac{\left[\frac{1}{1-\lambda^2}\sigma_\varepsilon^2 \left(\widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)'(\boldsymbol{\phi}_i \odot \mathbf{w}_i) + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)}\right)\right]\left[\frac{\lambda^{s-2}}{1+\lambda}\sigma_\varepsilon^2\left(1 + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)'(\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right)\right]}{\frac{\sigma_\alpha^2}{(1-\lambda)^2} + \frac{1}{1-\lambda^2}\sigma_\varepsilon^2\left(1 + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)'(\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right)} \\
& - \frac{\lambda^{s-2}}{1+\lambda}\sigma_\varepsilon^2\left[\widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)'(\boldsymbol{\phi}_i \odot \mathbf{w}_i) + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)}\right] \\
\neq \;& 0. \qquad (38)
\end{aligned}$$

[15] Breusch, Qian, Schmidt and Wyhowski (1999) provide a general treatment of redundancy of moment conditions.

In this case we have

$$\begin{aligned}
E\left[\Delta\boldsymbol{\upsilon}_i\Delta\boldsymbol{\upsilon}_i'\right] &= E\left[\left(\phi\left(\mathbf{W}\otimes\mathbf{I}_T\right)\Delta\boldsymbol{\varepsilon}_i+\Delta\boldsymbol{\varepsilon}_i\right)\left(\phi\left(\mathbf{W}\otimes\mathbf{I}_T\right)\Delta\boldsymbol{\varepsilon}_i+\Delta\boldsymbol{\varepsilon}_i\right)'\right] \\
&= E\left[\left[\phi\left(\mathbf{W}\otimes\mathbf{I}_T\right)+\mathbf{I}_{NT}\right]\Delta\boldsymbol{\varepsilon}_i\Delta\boldsymbol{\varepsilon}_i'\left[\phi\left(\mathbf{W}'\otimes\mathbf{I}_T\right)+\mathbf{I}_{NT}\right]\right] \\
&= \sigma_\varepsilon^2\left[\phi\left(\mathbf{W}\otimes\mathbf{I}_T\right)+\mathbf{I}_{NT}\right]\left(\mathbf{I}_N\otimes H_i\right)\left[\phi\left(\mathbf{W}'\otimes\mathbf{I}_T\right)+\mathbf{I}_{NT}\right] \\
&= \sigma_\varepsilon^2\left[\phi\left(\mathbf{W}\otimes\mathbf{I}_T\right)\left(\mathbf{I}_N\otimes\mathbf{H}_i\right)+\mathbf{I}_{NT}\left(\mathbf{I}_N\otimes\mathbf{H}_i\right)\right]\left[\phi\left(\mathbf{W}'\otimes\mathbf{I}_T\right)+\mathbf{I}_{NT}\right] \\
&= \sigma_\varepsilon^2\phi\left(\mathbf{W}\otimes\mathbf{H}_i\right)\phi\left(\mathbf{W}'\otimes\mathbf{I}_T\right)+\sigma_\varepsilon^2\phi\left(\mathbf{W}\otimes\mathbf{H}_i\right) \\
&\quad +\sigma_\varepsilon^2\left(\mathbf{I}_N\otimes\mathbf{H}_i\right)\phi\left(\mathbf{W}'\otimes\mathbf{I}_T\right)+\sigma_\varepsilon^2\left(\mathbf{I}_N\otimes\mathbf{H}_i\right) \\
&= \sigma_\varepsilon^2\left[\phi^2\mathbf{W}\mathbf{W}'+\phi\left(\mathbf{W}+\mathbf{W}'\right)+\mathbf{I}_N\right]\otimes\mathbf{H}_i \qquad (39)
\end{aligned}$$

and therefore $\left[\phi^2\mathbf{W}\mathbf{W}'+\phi\left(\mathbf{W}+\mathbf{W}'\right)+\mathbf{I}_N\right]$ replaces $\mathbf{I}_N$ in the expression for the weighting matrix of DIF GMM in (20). A similar point applies to $\mathbf{I}_N$ in (28) for SYS GMM. Of course, in practice ϕ is unknown; one option is to replace ϕ with an arbitrary value (say $\phi = 0.5$) at first stage, and then obtain an estimate of ϕ by solving the following quadratic equation:

$$\widehat{\phi}_t^2 r_t(1) - \widehat{\phi}_t + r_t(1) = 0 \qquad (40)$$

where $r_t(1) = est.corr.\left(\widehat{\upsilon}_{i,t}, \widehat{\upsilon}_{j,t}\right)$ and $\widehat{\upsilon}_{i,t}$ is the first-stage residual of unit i for $t = 2, ..., T$. (40) has two solutions for each t, but given that $r_t^{-1}(1) = \widehat{\phi}_t + \widehat{\phi}_t^{-1}$ one root is the reciprocal of the other, which implies that the estimator for ϕ at time t equals

$$\widehat{\phi}_t = \frac{1-\sqrt{1-4r_t^2(1)}}{2r_t(1)} \qquad (41)$$

The other solution can be ruled out since it will have an absolute value greater than one, which is not possible given the restriction $|\phi| < 1$. A simple average $\widehat{\phi} = \frac{1}{T}\sum\widehat{\phi}_t$ can then be constructed to provide an estimate of $\widehat{\phi}$.

6 Consistent GMM Estimation under both Spatial and Factor Error Structure

The moment conditions analysed in the previous section can be particularly useful in general error processes that include unobserved common factors and spatial correlations. This is because while the standard moment conditions in (10) and (23) are invalidated in this case (see Sarafidis and Robertson, 2008), it turns out that the moment conditions obtained from those individuals which unit i is spatially correlated with, are still valid.[16]

As an illustration, consider a mixture case for the error term, where there is a spatial moving average process of first-order along with a K-factor structure. In this case the

[16] Notice that the use of other instruments with respect to individual i will not help either, unless these instruments are not functions of (lagged values of) y and certain regularity conditions hold true, such as those in Sarafidis, Yamagata and Robertson (2008).

basic model can be written in terms of deviations from time-specific averages as follows

$$\begin{aligned} \underline{y}_{i,t} &= \lambda \underline{y}_{i,t-1} + \underline{\upsilon}_{i,t}, \\ \underline{\upsilon}_{i,t} &= \underline{a}_i + \underline{u}_{i,t}, \quad \underline{u}_{i,t} = \sum_{m=1}^{K} \underline{\phi}_{i,m} f_{m,t} + \underline{\varepsilon}_{i,t} + \theta \underline{\varepsilon}_{j,t}, \end{aligned} \tag{42}$$

where $j = i \,(\mathrm{mod}\, N) + 1$ and θ denotes the SMA(1) coefficient in order to make the distinction between the factor loadings and the spatial parameter more clear. A similar error process that is subject to both spatial correlations and common unobserved factors is studied by Pesaran and Tosetti (2007) for T large. Notice that it is not important what the particular form of spatial dependence in $\varepsilon_{i,t}$ is $-$ what is required essentially is that there is a subset of $\mathbf{W}$ that satisfies conditions (ii) and (iii) of Theorem 2, as in (8).

The following proposition demonstrates an interesting result.

Proposition 8 *Under Assumptions 1-6, the panel autoregressive model in* (42) *can be estimated consistently using method of moments estimators that rely on the following moment conditions*

Moment Conditions for DIF GMM:

$$\mathrm{plim}_{N\to\infty} \frac{1}{N} \left[\sum_{i=1}^{N} \sum_{t=s+1}^{T} \underline{y}_{j,t-s} \Delta \underline{\upsilon}_{i,t} \right] = 0; \text{ for } s = 2, ..., T-1, \tag{43}$$

and

$$\mathrm{plim}_{N\to\infty} \frac{1}{N} \left[\sum_{i=1}^{N} \sum_{t=s+1}^{T} \underline{y}_{j,t-s} \Delta \underline{y}_{i,t-1} \right] = -\theta \frac{(T-s)}{1+\lambda} \sigma_{\varepsilon}^2. \tag{44}$$

Moment Conditions for SYS GMM:

$$\mathrm{plim}_{N\to\infty} \frac{1}{N} \left[\sum_{i=1}^{N} \sum_{t=3}^{T} \Delta \underline{y}_{j,t-1} \underline{\upsilon}_{i,t} \right] = 0 \tag{45}$$

and

$$\mathrm{plim}_{N\to\infty} \frac{1}{N} \left[\sum_{i=1}^{N} \sum_{t=3}^{T} \Delta \underline{y}_{j,t-1} \underline{y}_{i,t-1} \right] = \theta \frac{(T-2)}{1+\lambda} \sigma_{\varepsilon}^2. \tag{46}$$

Proof. See Appendix C. ■

The above implies that the model given in (42) can be estimated consistently using a simple IV estimator that employs $\underline{y}_{j,t-2}$ as an instrument for $\Delta \underline{y}_{i,t-1}$, or a first-differenced GMM estimator that instruments $\Delta \underline{y}_{i,t-1}$ by $\underline{y}_{j,t-s}$ for $s = 2, 3, ...$, and a system GMM estimator that uses $\Delta \underline{y}_{j,t-1}$ as an instrument for $\underline{y}_{i,t-1}$ in the levels equations. This is because the correlation between $\underline{y}_{j,t-s}$and $\Delta \underline{y}_{i,t-1}$ (or between $\Delta \underline{y}_{j,t-1}$ and $\underline{y}_{i,t-1}$ in levels) is non-zero while the correlation between $\underline{y}_{j,t-s}$ and $\Delta \underline{\upsilon}_{i,t}$ (and $\Delta \underline{y}_{j,t-1}$ and $\underline{\upsilon}_{i,t}$ in levels) remains zero. Therefore, defining $\mathbf{Z}_{MM} = \left(\mathbf{Z}_1^{MM}, ..., \mathbf{Z}_N^{MM}\right)'$ with $\mathbf{Z}_i^{MM} = \left(\underline{y}_{j,1}, \underline{y}_{j,2}, ..., \underline{y}_{j,T-2}\right)$, as well as the following matrices of instruments

$$\underline{\mathbf{Z}}_i^{\dagger} = \begin{bmatrix} \underline{y}_{j,1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \underline{y}_{j,1} & \underline{y}_{j,2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \underline{y}_{j,1} & \underline{y}_{j,2} & \cdots & \underline{y}_{j,T-2} \end{bmatrix} ; \; \Delta \underline{\boldsymbol{\upsilon}}_i = \begin{bmatrix} \Delta \underline{\upsilon}_{i,3} \\ \Delta \underline{\upsilon}_{i,4} \\ \vdots \\ \Delta \underline{\upsilon}_{i,T} \end{bmatrix}, \quad (47)$$

and

$$\underline{\mathbf{Z}}_i^{\dagger sys} = \begin{bmatrix} \underline{\mathbf{Z}}_i^{\dagger} & 0 & \cdots & 0 \\ 0 & \Delta \underline{y}_{j,2} & & 0 \\ . & . & \ddots & . \\ 0 & 0 & \cdots & \Delta \underline{y}_{j,T} \end{bmatrix} ; \; \underline{\boldsymbol{\upsilon}}_i^{sys} = \begin{bmatrix} \Delta \underline{\boldsymbol{\upsilon}}_i \\ \underline{\boldsymbol{\upsilon}}_i \end{bmatrix}, \quad (48)$$

Proposition 8 implies that the following moment estimators are valid.

$$\widehat{\lambda}_{IV}^{\dagger} = \left(\mathbf{Z}'_{MM} \Delta \underline{\mathbf{y}}_{-1}\right)^{-1} \left(\mathbf{Z}'_{MM} \Delta \underline{\mathbf{y}}\right), \quad (49)$$

$$\widehat{\lambda}_{DIF\ GMM}^{\dagger} = \left[\frac{1}{N} \Delta \underline{\mathbf{y}}'_{-1} \underline{\mathbf{Z}} \left(\widehat{\mathbf{A}}_N^{\dagger}\right)^{-1} \underline{\mathbf{Z}}' \Delta \underline{\mathbf{y}}_{-1}\right]^{-1} \left[\frac{1}{N} \Delta \underline{\mathbf{y}}'_{-1} \underline{\mathbf{Z}} \left(\widehat{\mathbf{A}}_N^{\dagger}\right)^{-1} \underline{\mathbf{Z}}' \Delta \underline{\mathbf{y}}\right], \quad (50)$$

and

$$\widehat{\lambda}_{SYS\ GMM}^{\dagger} = \left[\frac{1}{N} \underline{\mathbf{Y}}'_{-1} \underline{\mathbf{Z}}^{sys} \left(\widehat{\mathbf{A}}_{\mathbf{N,sys}}^{\dagger}\right)^{-1} \underline{\mathbf{Z}}^{sys\prime} \underline{\mathbf{Y}}_{-1}\right]^{-1} \left[\frac{1}{N} \underline{\mathbf{Y}}'_{-1} \underline{\mathbf{Z}}^{sys} \left(\widehat{\mathbf{A}}_{\mathbf{N,sys}}^{\dagger}\right)^{-1} \underline{\mathbf{Z}}^{sys\prime} \underline{\mathbf{Y}}\right]. \quad (51)$$

$\left(\widehat{\mathbf{A}}_N^{\dagger}\right)^{-1}$ is the weighting matrix of the two-step first-differenced GMM estimator, which can be estimated from

$$\widehat{\mathbf{A}}_N^{\dagger} = \underline{\mathbf{Z}}' \left[\left(\Delta \widehat{\underline{\boldsymbol{\upsilon}}} \Delta \widehat{\underline{\boldsymbol{\upsilon}}}'\right) \otimes \mathbf{H}_i\right] \underline{\mathbf{Z}}, \quad (52)$$

where $\mathbf{H}_i$ has been defined in (21) and $\Delta \widehat{\underline{\boldsymbol{\upsilon}}}$ is an $N \times (T-2)$ matrix of residuals, obtained from a first-step first-differenced GMM estimator. Notice that the least-squares estimate of $\Delta \widehat{\underline{\boldsymbol{\upsilon}}} \Delta \widehat{\underline{\boldsymbol{\upsilon}}}'$ in (52) is rank deficient because it is an $N \times N$ matrix and has rank $T-2$. The matrix inside the square brackets of (52) is also rank deficient because it is a square matrix of order $N(T-2)$ and has rank $(T-2)^2$. However, $\widehat{\mathbf{A}}_N^{\dagger}$ is a square $\zeta \times \zeta$ matrix, which has rank equal to min $\left(\zeta, (T-2)^2\right)$. Therefore, provided that we do not use too many instruments, i.e. $\zeta \leq (T-2)^2$, (52) will be of full rank and the weighting matrix will exist.

$\left(\widehat{\mathbf{A}}^{\dagger}_{N,sys}\right)^{-1}$ is the weighting matrix of the two-step system GMM estimator, which can be estimated from

$$\widehat{\mathbf{A}}^{\dagger}_{N,sys} = \underline{\mathbf{Z}}^{sys\prime}\left(\widehat{\mathbf{Q}}\right)\underline{\mathbf{Z}}^{sys}, \tag{53}$$

with $\widehat{\mathbf{Q}}$ being equal to

$$\widehat{\mathbf{Q}} = \begin{bmatrix} \left(\Delta\widehat{\underline{\boldsymbol{v}}}\Delta\widehat{\underline{\boldsymbol{v}}}'\right)\otimes(H_i) & \mathbf{0} \\ \mathbf{0} & \left(\widehat{\underline{\boldsymbol{v}}}\widehat{\underline{\boldsymbol{v}}}'\right)\otimes(I_{T-2}) \end{bmatrix}. \tag{54}$$

7 Properties of GMM Estimators

To investigate the properties of these moment estimators we follow the approach by Blundell and Bond (1998) and we consider the case where $T = 3$, for which there is only a single instrument available for the endogenous regressor, both in the first-differenced equations and those in levels. In this way, DIF GMM and SYS GMM reduce to simple instrumental variable estimators and the corresponding first-stage regressions may help to analyse the 'strength' of the instruments used as a function of the parameters of interest in more general cases.

7.1 Equations in First-Differences

For the equations in first-differences, the first-stage regression is given by

$$\Delta\underline{y}_{i,2} = \pi^d \underline{y}_{j,1} + w_i \tag{55}$$

where w_i is an error term. The least-squares estimator for π^d, which reflects the strength of the relation between the instrument and the endogenous regressor, is equal to

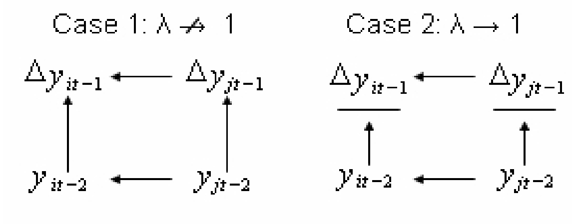
$$\widehat{\pi}^d = \frac{\sum_{i=1}^N \underline{y}_{j,1}\Delta\underline{y}_{i,2}}{\sum_{i=1}^N \underline{y}_{j,1}^2}. \tag{56}$$

Using Assumptions 1-6 in model (42) it is straightforward to show that the plim of $\widehat{\pi}^d$ equals

$$\begin{aligned}
&\operatorname{plim}_{N\to\infty}\left(\widehat{\pi}^d\right) \\
=\;& \frac{(\lambda-1)\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \underline{y}_{j,1}\underline{y}_{i,1} + \operatorname{plim}_{N\to\infty}\frac{1}{N}\sum \underline{y}_{j,1}\left(\underline{y}_{i,2}-\lambda\underline{y}_{i,1}\right)}{\operatorname{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \underline{y}_{j,1}^2} \\
=\;& (\lambda-1)\,\theta\frac{\sigma_\varepsilon^2}{1-\lambda^2}\left[\frac{\sigma_\alpha^2}{(1-\lambda)^2}+\frac{\sigma_\varepsilon^2}{1-\lambda^2}\left(1+\theta^2\right)+\mathbf{w}_1'\boldsymbol{\Sigma}_\phi\mathbf{w}_1\right]^{-1} \\
=\;& (\lambda-1)\,\theta\left[\frac{\kappa}{\sigma_\alpha^2/\sigma_\varepsilon^2+\left(1+\theta^2\right)\kappa+\left(1-\lambda^2\right)\kappa\left[\mathbf{w}_1'\left(\boldsymbol{\Sigma}_\phi/\sigma_\varepsilon^2\right)\mathbf{w}_1\right]}\right],
\end{aligned} \tag{57}$$

where $\kappa = \frac{1-\lambda}{1+\lambda}$ and $\mathbf{w}_1 = \sum_{s=0}^{\infty} \lambda^s \mathbf{f}_{1-s}$.

Thus, we can see that for fixed T the plim of $\widehat{\pi}^d$ depends on various parameters, namely λ, θ, σ_α^2, $\mathbf{\Sigma}_\phi$ and σ_ε^2. For example, as $\lambda \to 1$ the plim of the estimator converges to zero, which implies that the correlation between $\underline{y}_{j,t-2}$ and $\Delta\underline{y}_{i,t-1}$ becomes weak. The intuition behind this is illustrated in the following figure, which shows two cases of λ when the values of σ_α^2, $\mathbf{\Sigma}_\phi$, σ_ε^2 and θ are held fixed:



Weak instruments with cross section dependence.

When $\lambda \nrightarrow 1$, $\underline{y}_{j,1}$ is correlated with $\Delta\underline{y}_{i,2}$ and since $Cov\left(\underline{y}_{j,1}, \Delta\underline{\upsilon}_{i,3}\right) = 0$ $\underline{y}_{j,1}$ is a valid instrument. Note that the use of $\underline{y}_{i,1}$ as an instrument is not valid here because $\mathbf{\Sigma}_\phi \neq 0$ and therefore $Cov\left(\underline{y}_{i,1}, \Delta\underline{\upsilon}_{i,3}\right) \neq 0$. On the other hand, as $\lambda \to 1$ the correlation between $\underline{y}_{j,1}$ and $\Delta\underline{y}_{j,2}$ becomes weak; this is because the link between $\Delta\underline{y}_{j,2}$ and $\Delta\underline{y}_{i,2}$ is not effective anymore since $\underline{y}_{j,1}$ is poorly correlated with $\Delta\underline{y}_{j,2}$, while the link between $\underline{y}_{j,1}$ and $\underline{y}_{i,1}$ does not help either because $\underline{y}_{i,1}$ is poorly correlated with $\Delta\underline{y}_{i,2}$.

When there is no variation in the factor loadings across i, $\mathbf{\Sigma}_\phi = 0$ and the plim of $\widehat{\pi}^d$ remains non-zero but of course in this case $\underline{y}_{i,1}$ is also valid as instrument. On the other hand, for a given non-zero value of θ and $|\lambda| < 1$ the plim of the estimator converges to zero as either $\left(\sigma_\alpha^2/\sigma_\varepsilon^2\right) \to \infty$ or $\left(\mathbf{\Sigma}_\phi/\sigma_\varepsilon^2\right) \to \infty$. The former result is similar to Blundell and Bond (1998). Interestingly, the same appears to apply for the ratio between $\mathbf{\Sigma}_\phi$ and σ_ε^2. Intuitively, this is because the contribution of the spatial component of the error process in $\Delta\underline{\upsilon}_{i,3}$ (and thereby the correlation between Δy_{i2} and y_{j1}) diminishes with high values of $\mathbf{\Sigma}_\phi$ and increases with high values of σ_ε^2.

7.2 Equations in Levels

For the equations in levels the first-stage regression is given by

$$\underline{y}_{i,2} = \pi^l \Delta\underline{y}_{j,2} + w_i^l \tag{58}$$

and the least squares estimator of π^l equals

$$\widehat{\pi}^l = \frac{\sum_{i=1}^N \Delta\underline{y}_{j,2}\underline{y}_{i,2}}{\sum_{i=1}^N \left(\Delta\underline{y}_{j,2}\right)^2} \tag{59}$$

Using Assumptions 1-4, it is straightforward to show that the plim of $\widehat{\pi}^l$ equals

$$
\begin{aligned}
& \text{plim}_{N\to\infty}\left(\widehat{\pi}^l\right) \\
= \; & \frac{\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \underline{y}_{j,2}\underline{y}_{i,2} - \lambda\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \underline{y}_{j,1}\underline{y}_{i,1} - \text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \underline{y}_{j,1}\underline{\upsilon}_{i,2}}{\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\left(\Delta\underline{y}_{j,2}\right)^2} \\
= \; & \frac{\theta}{2\left(1+\phi^2\right)+(1+\lambda)\left[\Delta\mathbf{w}_2'\left(\boldsymbol{\Sigma}_\phi/\sigma_\varepsilon^2\right)\boldsymbol{\Delta}\mathbf{w}_2\right]} \qquad (60)
\end{aligned}
$$

where $\Delta\mathbf{w}_2 = \sum_{\tau=0}^{\infty}\lambda^\tau\Delta\mathbf{f}_{2-\tau}$. Here we can see that as $\lambda \to 1$ the above expression converges to

$$\text{plim}_{N\to\infty}\left(\widehat{\pi}^d\right) = \frac{1}{2}\frac{\theta}{\left(1+\phi^2\right)+\Delta\mathbf{w}_2'\left(\boldsymbol{\Sigma}_\phi/\sigma_\varepsilon^2\right)\boldsymbol{\Delta}\mathbf{w}_2} \qquad (61)$$

and so $\Delta\underline{y}_{j,2}$ remains informative as an instrument for $\underline{y}_{i,2}$, provided of course that $\theta \neq 0$. In addition, when $\boldsymbol{\Sigma}_\phi = 0$ the random element in (61) disappears and the expression becomes equal to a constant number − specifically, $\text{plim}_{N\to\infty}\left(\widehat{\pi}^l\right) = \theta/\left[2\left(1+\theta^2\right)\right]$. Similarly to (57), the plim of $\widehat{\pi}^l$ converges to zero as $\left(\boldsymbol{\Sigma}_\phi/\sigma_\varepsilon^2\right) \to \infty$ for the same reason that has been discussed previously − that is, because the contribution of the spatial component of the error process in υ_{i3} diminishes.

8 Small Sample Properties of Moment Estimators

This section investigates the finite-sample performance of the various estimators proposed in this paper using simulated data. The main focus of the analysis lies on the impact of the relative importance of the unobserved factors in the total error process for different values of N, T and λ.

8.1 Monte Carlo Design

The underlying data generating process is given by

$$
\begin{aligned}
y_{it} &= \lambda y_{it-1} + \alpha_i + u_{it}, \\
u_{it} &= \phi_i f_t + \varepsilon_{it} + \theta\varepsilon_{jt},\; i = 1,2,...,N;\, t = -48,-47,...,T. \qquad (62)
\end{aligned}
$$

where $\alpha_i \sim iidN\left(0,\sigma_\alpha^2\right)$, $\varepsilon_{it} \sim iidN\left(0,\sigma_\varepsilon^2\right)$, $f_t \sim iidN\left(0,\sigma_f^2\right)$ and $j = i\,(\text{mod}\,N)$. Also, the factor loadings are drawn from

$$\phi_i \sim iidU\left[-0.25, 0.25\right] \qquad (63)$$

The performance of GMM estimation depends crucially upon the ratio of the two variance components, a_i and u_{it}, on $var(y_{it})$ as shown in (57). This implies that as the value of λ increases, or the amount of cross-sectional dependence decreases, the impact of α_i on $var(y_{it})$ will tend to become larger and thereby comparisons across experiments with different levels of cross-sectional dependence will not be valid. To control this ratio we use the following simple result

$$var(y_{it}) = var\left[\frac{\alpha_i}{1-\lambda} + \left(\sum_{s=0}^{\infty} \lambda^s u_{it-s}\right)\right] = \frac{\sigma_\alpha^2}{(1-\lambda)^2} + \frac{\sigma_u^2}{1-\lambda^2} \tag{64}$$

and we set $\sigma_\alpha^2 = \psi\left[(1-\lambda)/(1+\lambda)\right]\sigma_u^2$ with $\psi = 1$[17].

In addition to $\sigma_\alpha^2/\sigma_u^2$, the performance of the estimators will depend on the proportion of σ_u^2 attributed to the factor structure in u_{it} − hereafter this proportion is denoted by $\zeta_{(d)}$, $d = 1, ..., 4$. Therefore, noticing that

$$\sigma_u^2 = \left(\mu_\phi\right)^2 \sigma_f^2 + \sigma_\phi^2 \sigma_f^2 + \sigma_\varepsilon^2\left(1+\theta^2\right) \tag{65}$$

and normalising $\sigma_f^2 = 1$, we can produce the following result

$$\sigma_\varepsilon^2 = \frac{\left(1-\xi_{(d)}\right)}{\xi_{(d)}} \frac{\left(\mu_\phi\right)^2 + \sigma_\phi^2}{\left(1+\theta^2\right)} \tag{66}$$

Since the values of $\left(\mu_\phi\right)^2$ and σ_ϕ^2 are determined solely by (63) and so they are fixed, normalising $\theta = 0.5$ implies that σ_ε^2 will change only according to $\xi_{(d)}$. As this ratio increases, the impact of the factor structure in the error process will rise. We choose the following values for $\xi_{(d)}$:

$$\begin{cases} \text{Low impact of factor structure on } u_{it}\text{:} & \xi_{(1)} = 1/3 \\ \text{Medium impact of factor structure on } u_{it}\text{:} & \xi_{(2)} = 1/2 \\ \text{Medium-to-high impact of factor structure on } u_{it}\text{:} & \xi_{(3)} = 2/3 \\ \text{High impact of factor structure on } u_{it}\text{:} & \xi_{(4)} = 3/4 \end{cases}$$

We consider $N = 400$, 800 and $T = 6$, 10, since our focus is T fixed, $N \to \infty$. λ alternates between 0.5, 0.7 and 0.9. The initial value of y_{it} has been set equal to zero but the first 50 observations have been discarded before choosing the sample, so as to ensure that the initial zero values do not have an impact on the results. All experiments are based on 2,000 replications.

8.2 Results

Since the IV estimator has no finite moments, Tables A1-A2 in the appendix report median bias and median square error for all estimators. FE is the fixed effects estimator, IV is the simple instrumental variables estimator that uses y_{it-2} as a single instrument

[17] See Kiviet (1995) and Bun and Kiviet (2006).

for Δy_{it-1} and DIF and SYS denote the first-differenced and system GMM estimators respectively[18]. The superscript '*' indicates that the corresponding estimator uses instrument(s) with respect to another cross-sectional, unit j.

As expected, the performance of all estimators depends on $\xi_{(d)}$, the value of λ and the size of T and N. Specifically, as the value of ξ increases for a given value of λ, T and N, the estimators suffer a rise in bias and in RMSE. This is natural because as the relative impact of the factor structure in the total error process increases, the invalidity of the instruments used with respect to unit i itself (such as in IV, DIF and SYS) is magnified. For the estimators that make use of instruments with respect to unit j, the rise in bias and RMSE is also intuitive because as ξ increases, the contribution of the spatial component in the error process − and thereby the correlation between Δy_{it-1} and y_{jt-2} − diminishes.

Having said that, two things are clear from these results; first, IV^*, DIF^* and SYS^* outperform IV, DIF and SYS respectively under all circumstances. Second, the relative performance of IV^*, DIF^* and SYS^* improves with larger values of ξ. This is also intuitive − ultimately, as $\xi \to 0$ the factor structure in the error process diminishes and the asymptotic bias of IV, DIF and SYS approaches zero. Notice also that in terms of RMSE, SYS* performs better than DIF*, which performs better than IV*, with the relative difference in performance being increased according to the value of λ. As T rises, the performance of the estimators improves without exception.

Finally, it is important to emphasise that as the size of N increases, the bias and RMSE of IV^*, DIF^* and SYS^* decreases considerably. This is not the case for the conventional estimators, IV, DIF and SYS, the performance of which − if anything − deteriorates with larger values of N.

9 Concluding Remarks

Error cross-sectional dependence is an increasingly popular research topic in the analysis of panel data. Despite this fact, the issue has not attracted much attention in GMM estimation of short dynamic panels, where it is commonly assumed that the regression errors are independent across cross-sectional individuals. This paper has shown that, in fact, independence or uncorrelatedness is not necessary for GMM consistency or asymptotic efficiency − rather, it is sufficient that, if there is such correlation in the errors, this is weak in the sense that the sequence of the covariances of the disturbances across individuals i and j at time t, given the conditioning set of all time-invariant characteristics of individuals i and j, converges absolutely as $N \to \infty$. If this condition is not satisfied, the errors are said to be strongly correlated. Spatial dependence presents an example of cross-sectionally weakly correlated errors while the factor structure dependence provides an example of cross-sectionally strongly correlated errors. As a result, the standard dynamic panel GMM estimators that exist in the literature remain con-

[18] DIF and SYS are estimated in two steps and they use y_{it-2} and y_{it-3} as instruments for Δy_{it-1} in the first-differenced equations. Furthermore, SYS GMM uses the optimal weighting matrix (when $\sigma^2_\alpha = 0$), as derived in Windmeijer (2000).

sistent under spatially correlated errors but not so under a factor structure. When the errors are cross-sectionally weakly correlated there are additional moment conditions that arise − in particular, instruments with respect to the individual(s) which unit i is correlated with. We demonstrate that these moment conditions are particularly useful when the errors are subject to both weak and strong correlations, in which case the standard instruments are invalid. The properties of the resulting GMM estimators have been analysed under different circumstances. Simulated experiments have shown that these estimators outperform the conventional ones, in terms of both median bias and root median square error. This result is magnified as the impact of the factor structure in the total error process increases. In addition, larger values of N are accompanied by a considerable decrease in bias and RMSE for the estimators put forward in this paper. This is not the case with the conventional estimators, the performance of which is naturally not affected by the size of N.

References

[1] Ahn, S.C and P. Schmidt, (1995) "Efficient Estimations of Models for Dynamic Panel Data", *Journal of Econometrics*, 68, 5-28.

[2] Anderson, T.W. and C. Hsiao (1981) "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, 76, 598-606.

[3] Arellano, M. (1993) "On the Testing of Correlated Effects with Panel Data", *Journal of Econometrics*, 59, 87-97.

[4] Arellano, M. and S. Bond (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58, 277-297.

[5] Arellano, M. and O. Bover (1995) "Another Look at the Instrumental Variable Estimation of Error-Component Models", *Journal of Econometrics*, 68, 29-51.

[6] Bai, J. (2005) "Panel Data Models with Interactive Fixed Effects", unpublished manuscript.

[7] Baltagi, B., Bresson, G. and A. Pirotte (2007) "Panel Unit Root Tests and Spatial Dependence", *Journal of Applied Econometrics*, 22, 339-360.

[8] Blundell, R. and S. Bond (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics*, 87, 115-143.

[9] Bover, O. and N. Watson (2005) "Are there economies of scale in the demand for money by firms? Some panel data estimates", *Journal of Monetary Economics*, 52, 1569-1589.

[10] Breusch, T., Qian, H., Schmidt, P. and D. Wyhowski, 1999, Redundancy of moment conditions. Journal of Econometrics 91, 89-111.

[11] Bun, M.J.G. and J.F. Kiviet (2006) "The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models", *Journal of Econometrics* 132, 409-444.

[12] Coakley, J., Fuertes, A. and Smith, R. (2002) "A Principal Components Approach to Cross-Section Dependence in Panels", Unpublished manuscript, Birckbeck College, University of London.

[13] Conley, T.G. (1999) "GMM Estimation with cross-sectionalal Dependence", *Journal of Econometrics,* 92, 1-45.

[14] Goldberger, A.S. (1972) "Structural Equation Methods in the Social Sciences", *Econometrica*, 40, 979-1001.

[15] Jöreskog, K.G. and A.S. Goldberger (1975) "Estimation of a model with multiple indicators and multiple causes of a single latent variable" *Journal of the American Statistical Association*, 70, 631-639.

[16] Kapoor, M., Kelejian, H., and I.R. Prucha (2007) "Panel Data Models with Spatially Correlated Error Components", *Journal of Econometrics* 140, 97-130.

[17] Kelejian, H and D. Robinson (1995) "Spatial Correlation: A Suggested Alternative to the Autoregressive Model, in Anselin, L. and R.J. Florax (ed.), *New Directions in Spatial Econometrics*, 75-95, Springer-Verlag, Berlin.

[18] Kiviet, J.F., (1995) "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models", *Journal of Econometrics*, 68, 53-78.

[19] Lee, L. (2007) "GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models", *Journal of Econometrics* 137, 489-514.

[20] Moon, H.R. and B. Perron (2004) "Efficient Estimation of the SUR Cointegrating Regression Model and Testing for Purchasing Power Parity", *Econometric Reviews*, Vol. 23, 293-323.

[21] Pesaran, H. (2006) "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure", *Econometrica*, 74, 967-1012.

[22] Pesaran, H. and E. Tosetti (2007) "Large Panels with Common Factors and Spatial Correlations", unpublished manuscript.

[23] Phillips, P. and D. Sul (2003) "Dynamic Panel Estimation and Homogeneity Testing under cross-sectional Dependence", *The Econometrics Journal*, Vol. 6, 217-259.

[24] Phillips, P. and D. Sul (2007). "Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and cross-sectional Dependence", *Journal of Econometrics*, 137, 162-188.

[25] Robertson, D. and Symons, J. (2000) "Factor Residuals in SUR Regressions: Estimating Panels Allowing for cross-sectionalal Correlation", unpublished manuscript, Faculty of Economics and Politics, University of Cambridge.

[26] Sarafidis, V. Yamagata, T. and D. Robertson (2008) "A Test of error cross-sectional Dependence for a Linear Dynamic Panel Model with Regressors", unpublished manuscript.

[27] Sarafidis, V. and D. Robertson (2008) "On the Impact of cross-sectional Dependence in Short Dynamic Panel Estimation", unpublished manuscript.

[28] Stetzer, F. (1982) "Specifying Weights in Spatial Forecasting Models: The Results of Some Experiments", *Environment and Planning A*, 14, 571-584.

[29] Stout, W. (1974) "Almost Sure Convergence", Academic Press, New York.

[30] White, H. (2001) "Asymptotic Theory for Econometricians", Academic Press, New York.

[31] Windmeijer, F. (2000) "Efficiency Comparisons for a System GMM Estimator in Dynamic Panel Data Models", in R.D.H. Heijmans, D.S.G. Pollock and A. Satora (ed.), *Innovations in Multivariate Statistical Analysis*, Dordrecht: Kluwer Academic Publishers.

Appendices

A Proof of Theorem 2

The error process is given by $\upsilon_{i,t} = \alpha_i + \sum_{m=1}^{M} \phi_{i,m} w_{i,m} f_{m,t} + \varepsilon_{i,t}$. We maintain Assumptions 1-6 throughout and we also require that $E\left(\alpha_i \alpha_{i+k}\right) = 0$ for all $k \neq 0$. Notice that we do not wish to restrict M to be finite necessarily because spatial dependence processes arise from the above error structure by letting M grow with N. First, for $\upsilon_{i,t}$ to have finite variance we must have the following conditions hold true: (i) $w_{i,m} = O(1)\ \forall\ i,\ m$ and (ii) $R_i/N \rightarrow 1$ for $i = 1, ..., N$, where R_i is the number of elements in the i^{th} row of $\mathbf{W}$ that are either zero, or at most of order $N^{-1/2-\eta}$ – an individual element is denoted as $w_{i,m} = O\left(N^{-1/2-\eta}\right)$ for some $\eta \geq 0$. Of course if $M = O(1)$, condition (i) is sufficient to ensure finite variance. If M grows with N, (ii) implies that $||\mathbf{W}||_{\infty} = \max_i \sum_{m=1}^{M} |w_{i,m}| = O\left(N^{1/2}\right)$ but not vice versa.

Now, the correlation coefficient between $\upsilon_{i,t}$ and $\upsilon_{j,t}$ is given by

$$\sigma_{ij,t} = \frac{Cov\left(\upsilon_{i,t}, \upsilon_{j,t}|\Upsilon_{ij}\right)}{\left[Var\left(\upsilon_{i,t}|\Upsilon_{ij}\right) Var\left(\upsilon_{j,t}\right)|\Upsilon_{ij}\right]^{1/2}} = \frac{E\left(\upsilon_{i,t}\upsilon_{j,t}|\Upsilon_{ij}\right)}{\left[E\left(\upsilon_{i,t}^2|\Upsilon_{ij}\right) E\left(\upsilon_{j,t}^2|\Upsilon_{ij}\right)\right]^{1/2}}$$

$$= \frac{\sum_{m=1}^{M} \phi_{i,m}\phi_{j,m} w_{i,m} w_{j,m} \sigma_{f_m}^2}{\left[\left(\sigma_\alpha^2 + \sum_{m=1}^{M} \phi_{i,m}^2 w_{i,m}^2 \sigma_{f_m}^2 + \sigma_\varepsilon^2\right)\left(\sigma_\alpha^2 + \sum_{m=1}^{M} \phi_{j,m}^2 w_{j,m}^2 \sigma_{f_m}^2 + \sigma_\varepsilon^2\right)\right]^{1/2}}.$$

The condition $\sum_{j \neq i} |\sigma_{ij,t}| < \infty$, is automatically satisfied when the number of non-zero values in the m^{th} column of $\mathbf{W}$ is bounded, for all m. On the other hand, if this is not the case what we require is that $C_m/N \rightarrow 1$ for $m = 1, ..., M$, where C_m is the number of elements in the m^{th} column of $\mathbf{W}$ that are either zero, or of smaller order than $N^{-1/2}$, denoted as $w_{i,m} = o\left(N^{-1/2}\right)$. This means that $||\mathbf{W}||_1 = \max_m \sum_{i=1}^{N} |w_{i,m}| = o\left(N^{1/2}\right)$ but not vice versa. Furthermore, it implies that $w_{i,m} w_{j,m} = o\left(N^{-1}\right)$ for $m = 1, ..., M$. As a result, we have $\sum_{j\neq i}\sum_{m=1}^{M} \left|\phi_{i,m}\phi_{j,m} w_{i,m} w_{j,m}\right| \sigma_{f_m}^2 = O_p(1)$ and $\sum_j \sum_{m=1}^{M} \left|\phi_{j,m}^2 w_{j,m}^2\right| \sigma_{f_m}^2 = O_p(1)\ \forall\ k$. Since both σ_α^2 and σ_ε^2 are bounded, it follows that $\sum_{j\neq i} |\sigma_{ij,t}| < \infty$.

B Proof of Theorem 7

In the context of a linear regression model a set of moment conditions, Z_2, is redundant given another set of moment conditions, Z_1, if the (net of Z_1) covariance between Z_2 and the set of instrumented regressors is zero (see Breusch, Qian, Schmidt and Wyhowski (1999)). In our set up, this traslates to showing that $cov\left[\widetilde{\mathbf{y}}_{i,t-s}, \left(\Delta y_{i,t-1} - \widehat{\pi}_1 y_{i,t-s}\right)\right] = cov\left[\widetilde{\mathbf{y}}_{i,t-s}, \left(\Delta y_{i,t-1} - \frac{\sum_i \Delta y_{i,t-1} y_{i,t-s}}{\sum_i y_{i,t-s}^2} y_{i,t-s}\right)\right] \neq 0$ and $cov\left[\Delta\widetilde{\mathbf{y}}_{i,t-1}, \left(y_{i,t-1} - \widehat{\pi}_2 \Delta y_{i,t-1}\right)\right] = cov\left[\Delta\widetilde{\mathbf{y}}_{i,t-1}, \left(y_{i,t-1} - \frac{\sum_i y_{i,t-1}\Delta y_{i,t-1}}{\sum_i \left(\Delta y_{i,t-1}\right)^2}\Delta y_{i,t-1}\right)\right] \neq 0$, where $\widehat{\pi}_1$ denotes the least squares coefficient in the regression of $\Delta y_{i,t-1}$ on $y_{i,t-s}$ and $\widehat{\pi}_2$ denotes the least squares coefficient in the regression of $y_{i,t-1}$ on $\Delta y_{i,t-1}$.

Initially, notice that

$$\widetilde{\mathbf{y}}_{i,t} = \lambda \widetilde{\mathbf{y}}_{i,t-1} + \widetilde{\boldsymbol{\alpha}}_i + (\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}}_i)'\mathbf{f}_t + \widetilde{\boldsymbol{\varepsilon}}_{i,t},$$

where $(\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}}_i)'$ denotes the matrix that contains the non-zero rows of $(\boldsymbol{\Phi} \odot \mathbf{W}) \odot (\mathbf{w}_i \mathbf{i}_N')$ excluding the i^{th} row itself, $\widetilde{\boldsymbol{\alpha}}_i$ denotes the vector that contains the non-zero elements of $\left[(\alpha_1, \alpha_2, ..., \alpha_N)' \odot \mathbf{w}_i\right]'$ excluding α_i itself, and similarly for $\widetilde{\boldsymbol{\varepsilon}}_{i,t}$.

For the spatial moment conditions in the first-differenced equations, we have

$$
\begin{aligned}
& E\left[\widetilde{\mathbf{y}}_{i,t-s}\Delta y_{i,t-1}\right] = E\left[\left(\frac{\widetilde{\boldsymbol{\alpha}}_i}{1-\lambda} + (\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})' \sum_{\tau=0}^{\infty} \lambda^{\tau} \mathbf{f}_{t-s-\tau} + \sum_{\tau=0}^{\infty} \lambda^{\tau} \widetilde{\boldsymbol{\varepsilon}}_{i,t-s-\tau}\right)\right. \\
& \left.\cdot\left((\boldsymbol{\phi}_i \odot \mathbf{w}_i)' \sum_{\tau=0}^{\infty} \lambda^{\tau} \Delta \mathbf{f}_{t-1-\tau} + \sum_{\tau=0}^{\infty} \lambda^{\tau} \Delta \varepsilon_{i,t-1-\tau}\right)\right] \\
= \; & (\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})' E\left[\sum_{\tau=0}^{\infty} \lambda^{\tau} \mathbf{f}_{t-s-\tau} \sum_{\tau=0}^{\infty} \lambda^{\tau} \Delta \mathbf{f}'_{t-1-\tau}\right] (\boldsymbol{\phi}_i \odot \mathbf{w}_i) + (\widetilde{\boldsymbol{\phi} \odot \mathbf{w}_i})' E\left[\sum_{\tau=0}^{\infty} \lambda^{\tau} \mathbf{f}_{t-s-\tau} \sum_{\tau=0}^{\infty} \lambda^{\tau} \Delta \varepsilon_{i,t-1-\tau}\right] \\
& + E\left[\sum_{\tau=0}^{\infty} \lambda^{\tau} \widetilde{\boldsymbol{\varepsilon}}_{i,t-s-\tau} \sum_{\tau=0}^{\infty} \lambda^{\tau} \Delta \mathbf{f}'_{t-1-\tau}\right] (\boldsymbol{\phi}_i \odot \mathbf{w}_i) + E\left[\sum_{\tau=0}^{\infty} \lambda^{\tau} \widetilde{\boldsymbol{\varepsilon}}_{i,t-s-\tau} \sum_{\tau=0}^{\infty} \lambda^{\tau} \Delta \varepsilon_{i,t-1-\tau}\right]. \qquad (67)
\end{aligned}
$$

For a SEC process, (67) becomes equal to

$$
E\left[\widetilde{\mathbf{y}}_{i,t-s}\Delta y_{i,t-1}\right] = -\frac{\lambda^{s-2}}{1+\lambda}\sigma_f^2 (\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})' (\boldsymbol{\phi}_i \odot \mathbf{w}_i). \qquad (68)
$$

For a SMA process, (67) becomes equal to

$$
\begin{aligned}
& E\left[\widetilde{\mathbf{y}}_{i,t-s}\Delta y_{i,t-1}\right] = \\
= \; & (\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})' E\left[\sum_{\tau=0}^{\infty} \lambda^{\tau} \boldsymbol{\varepsilon}_{t-s-\tau} \sum_{\tau=0}^{\infty} \lambda^{\tau} \Delta \boldsymbol{\varepsilon}'_{t-1-\tau}\right] (\boldsymbol{\phi}_i \odot \mathbf{w}_i) + E\left[\sum_{\tau=0}^{\infty} \lambda^{\tau} \widetilde{\boldsymbol{\varepsilon}}_{i,t-s-\tau} \sum_{\tau=0}^{\infty} \lambda^{\tau} \Delta \boldsymbol{\varepsilon}'_{t-1-\tau}\right] (\boldsymbol{\phi}_i \odot \mathbf{w}_i) \\
= \; & -\frac{\lambda^{s-2}}{1+\lambda}\sigma_\varepsilon^2 \left[(\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})' (\boldsymbol{\phi}_i \odot \mathbf{w}_i) + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right]. \qquad (69)
\end{aligned}
$$

In addition, the covariance between $\widetilde{\mathbf{y}}_{i,t-s}$ and $y_{i,t-s}$ equals

$$
\begin{aligned}
& E\left[\widetilde{\mathbf{y}}_{i,t-s} y_{i,t-s}\right] = E\left[\left(\frac{\widetilde{\boldsymbol{\alpha}}_i}{1-\lambda} + (\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})' \sum_{\tau=0}^{\infty} \lambda^{\tau} \mathbf{f}_{t-s-\tau} + \sum_{\tau=0}^{\infty} \lambda^{\tau} \widetilde{\boldsymbol{\varepsilon}}_{i,t-s-\tau}\right)\right. \\
& \left.\cdot\left(\frac{\alpha_i}{1-\lambda} + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)' \sum_{\tau=0}^{\infty} \lambda^{\tau} \mathbf{f}_{t-s-\tau} + \sum_{\tau=0}^{\infty} \lambda^{\tau} \varepsilon_{i,t-s-\tau}\right)\right] \\
= \; & (\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})' E\left[\sum_{\tau=0}^{\infty} \lambda^{\tau} \mathbf{f}_{t-s-\tau} \sum_{\tau=0}^{\infty} \lambda^{\tau} \mathbf{f}'_{t-s-\tau}\right] (\boldsymbol{\phi}_i \odot \mathbf{w}_i) + (\widetilde{\boldsymbol{\phi} \odot \mathbf{w}_i})' E\left[\sum_{\tau=0}^{\infty} \lambda^{\tau} \mathbf{f}_{t-s-\tau} \sum_{\tau=0}^{\infty} \lambda^{\tau} \varepsilon_{i,t-s-\tau}\right] \\
& + E\left[\sum_{\tau=0}^{\infty} \lambda^{\tau} \widetilde{\boldsymbol{\varepsilon}}_{i,t-s-\tau} \sum_{\tau=0}^{\infty} \lambda^{\tau} \mathbf{f}'_{t-s-\tau}\right] (\boldsymbol{\phi}_i \odot \mathbf{w}_i) + E\left[\sum_{\tau=0}^{\infty} \lambda^{\tau} \widetilde{\boldsymbol{\varepsilon}}_{i,t-s-\tau} \sum_{\tau=0}^{\infty} \lambda^{\tau} \varepsilon_{i,t-s-\tau}\right]. \qquad (70)
\end{aligned}
$$

For a SEC process, (70) becomes equal to

$$
E\left[\widetilde{\mathbf{y}}_{i,t-s} y_{i,t-s}\right] = \frac{1}{1-\lambda^2}\sigma_f^2 (\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})' (\boldsymbol{\phi}_i \odot \mathbf{w}_i). \qquad (71)
$$

For a SMA process, (70) becomes equal to

$$
\begin{aligned}
& E\left[\widetilde{\mathbf{y}}_{i,t-s} y_{i,t-s}\right] = \\
= \; & (\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})' E\left[\sum_{\tau=0}^{\infty} \lambda^{\tau} \boldsymbol{\varepsilon}_{t-s-\tau} \sum_{\tau=0}^{\infty} \lambda^{\tau} \boldsymbol{\varepsilon}'_{t-s-\tau}\right] (\boldsymbol{\phi}_i \odot \mathbf{w}_i) + E\left[\sum_{\tau=0}^{\infty} \lambda^{\tau} \widetilde{\boldsymbol{\varepsilon}}_{i,t-s-\tau} \sum_{\tau=0}^{\infty} \lambda^{\tau} \boldsymbol{\varepsilon}'_{t-s-\tau}\right] (\boldsymbol{\phi}_i \odot \mathbf{w}_i) \\
= \; & \frac{1}{1-\lambda^2}\sigma_\varepsilon^2 \left[(\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})' (\boldsymbol{\phi}_i \odot \mathbf{w}_i) + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right]. \qquad (72)
\end{aligned}
$$

Furthermore, we have

$$E\left[y_{i,t-s}\Delta y_{i,t-1}\right]=E\left[\left(\frac{\alpha_i}{1-\lambda}+(\boldsymbol{\phi}_i\odot\mathbf{w}_i)'\sum_{\tau=0}^{\infty}\lambda^{\tau}\mathbf{f}_{t-s-\tau}+\sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{i,t-s-\tau}\right)\right.$$

$$\left.\cdot\left((\boldsymbol{\phi}_i\odot\mathbf{w}_i)'\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\mathbf{f}_{t-1-\tau}+\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{i,t-1-\tau}\right)\right]$$

$$= (\boldsymbol{\phi}_i\odot\mathbf{w}_i)'E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\mathbf{f}_{t-s-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\mathbf{f}'_{t-1-\tau}\right](\boldsymbol{\phi}_i\odot\mathbf{w}_i)+(\boldsymbol{\phi}\odot\mathbf{w}_i)'E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\mathbf{f}_{t-s-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{i,t-1-\tau}\right]$$

$$+E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{i,t-s-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\mathbf{f}'_{t-1-\tau}\right](\boldsymbol{\phi}_i\odot\mathbf{w}_i)+E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{i,t-s-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{i,t-1-\tau}\right] \quad (73)$$

Hence (73) becomes

$$E\left[y_{i,t-s}\Delta y_{i,t-1}\right]=-\frac{\lambda^{s-2}}{1+\lambda}\sigma_f^2\left(\boldsymbol{\phi}_i\odot\mathbf{w}_i\right)'\left(\boldsymbol{\phi}_i\odot\mathbf{w}_i\right), \quad (74)$$

for a SEC process, and

$$E\left[y_{i,t-s}\Delta y_{i,t-1}\right]=$$

$$(\boldsymbol{\phi}_i\odot\mathbf{w}_i)'E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\boldsymbol{\varepsilon}_{t-s-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\boldsymbol{\varepsilon}'_{t-1-\tau}\right](\boldsymbol{\phi}_i\odot\mathbf{w}_i)+E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{i,t-s-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{i,t-1-\tau}\right]$$

$$= -\frac{\lambda^{s-2}}{1+\lambda}\sigma_{\varepsilon}^2\left[1+(\boldsymbol{\phi}_i\odot\mathbf{w}_i)'(\boldsymbol{\phi}_i\odot\mathbf{w}_i)\right], \quad (75)$$

for a SMA process.

Last,

$$E\left[y_{i,t-s}^2\right]=E\left[\left(\frac{\alpha_i}{1-\lambda}+(\boldsymbol{\phi}_i\odot\mathbf{w}_i)'\sum_{\tau=0}^{\infty}\lambda^{\tau}\mathbf{f}_{t-s-\tau}+\sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{i,t-s-\tau}\right)^2\right]$$

$$= \begin{cases}\frac{\sigma_\alpha^2}{(1-\lambda)^2}+\frac{1}{1-\lambda^2}\sigma_f^2\left(\boldsymbol{\phi}_i\odot\mathbf{w}_i\right)'\left(\boldsymbol{\phi}_i\odot\mathbf{w}_i\right)+\frac{1}{1-\lambda^2}\sigma_\varepsilon^2 & \text{for a SEC process, and}\\ \frac{\sigma_\alpha^2}{(1-\lambda)^2}+\frac{1}{1-\lambda^2}\sigma_\varepsilon^2\left[1+(\boldsymbol{\phi}_i\odot\mathbf{w}_i)'(\boldsymbol{\phi}_i\odot\mathbf{w}_i)\right] & \text{for a SMA process.}\end{cases} \quad (76)$$

Combining (67), (67), (73) and (76) we obtain the following expressions for a SEC and a SMA process, respectively

$$cov\left[\widetilde{\mathbf{y}}_{i,t-s},\left(\Delta y_{i,t-1}-\widehat{\pi}_1 y_{i,t-s}\right)\right]=cov\left[\widetilde{\mathbf{y}}_{i,t-s},\left(\Delta y_{i,t-1}-\frac{\sum_i\Delta y_{i,t-1}y_{i,t-s}}{\sum_i y_{i,t-s}^2}y_{i,t-s}\right)\right]$$

$$= cov\left(\widetilde{\mathbf{y}}_{i,t-s},\Delta y_{i,t-1}\right)-\frac{cov\left(\widetilde{\mathbf{y}}_{i,t-s},y_{i,t-s}\right)cov\left(\Delta y_{i,t-1},y_{i,t-s}\right)}{var\left(y_{i,t-s}^2\right)}$$

$$= -\frac{\lambda^{s-2}}{1+\lambda}\sigma_f^2\widetilde{(\boldsymbol{\phi}_i\odot\mathbf{w}_i)}'(\boldsymbol{\phi}_i\odot\mathbf{w}_i)+\frac{\left[\frac{1}{1-\lambda^2}\sigma_f^2\widetilde{(\boldsymbol{\phi}_i\odot\mathbf{w}_i)}'(\boldsymbol{\phi}_i\odot\mathbf{w}_i)\right]\left[\frac{\lambda^{s-2}}{1+\lambda}\sigma_f^2\left(\boldsymbol{\phi}_i\odot\mathbf{w}_i\right)'\left(\boldsymbol{\phi}_i\odot\mathbf{w}_i\right)\right]}{\frac{\sigma_\alpha^2}{(1-\lambda)^2}+\frac{1}{1-\lambda^2}\sigma_f^2\left(\boldsymbol{\phi}_i\odot\mathbf{w}_i\right)'\left(\boldsymbol{\phi}_i\odot\mathbf{w}_i\right)+\frac{1}{1-\lambda^2}\sigma_\varepsilon^2} \quad (77)$$

and

$$
\begin{aligned}
& cov\left[\widetilde{\mathbf{y}}_{i,t-s}, (\Delta y_{i,t-1} - \widehat{\pi}_1 y_{i,t-s})\right] = \\
= \;& \frac{\left[\frac{1}{1-\lambda^2}\sigma_\varepsilon^2\left(\widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'(\boldsymbol{\phi}_i \odot \mathbf{w}_i) + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right)\right]\left[\frac{\lambda^{s-2}}{1+\lambda}\sigma_\varepsilon^2\left(1 + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)'(\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right)\right]}{\frac{\sigma_\alpha^2}{(1-\lambda)^2} + \frac{1}{1-\lambda^2}\sigma_\varepsilon^2\left(1 + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)'(\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right)} \\
& -\frac{\lambda^{s-2}}{1+\lambda}\sigma_\varepsilon^2\left[\widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'(\boldsymbol{\phi}_i \odot \mathbf{w}_i) + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right] \\
\neq \;& 0. \qquad (78)
\end{aligned}
$$

For the spatial moment conditions in the equations in levels, we have

$$
\begin{aligned}
& E\left[\Delta\widetilde{\mathbf{y}}_{i,t-1} y_{i,t-1}\right] = E\Bigg[\left(\widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\mathbf{f}_{t-1-\tau} + \sum_{\tau=0}^{\infty}\lambda^\tau \Delta\widetilde{\boldsymbol{\varepsilon}}_{i,t-1-\tau}\right) \\
& \cdot\left(\frac{\alpha_i}{1-\lambda} + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)'\sum_{\tau=0}^{\infty}\lambda^\tau \mathbf{f}_{t-1-\tau} + \sum_{\tau=0}^{\infty}\lambda^\tau \varepsilon_{i,t-1-\tau}\right)\Bigg] \\
= \;& \widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'E\left[\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\mathbf{f}_{t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^\tau \mathbf{f}'_{t-1-\tau}\right](\boldsymbol{\phi}_i \odot \mathbf{w}_i) + \widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'E\left[\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\mathbf{f}_{t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^\tau \varepsilon_{i,t-1-\tau}\right] \\
& + E\left[\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\widetilde{\boldsymbol{\varepsilon}}_{i,t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^\tau \mathbf{f}'_{t-1-\tau}\right](\boldsymbol{\phi}_i \odot \mathbf{w}_i) + E\left[\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\widetilde{\boldsymbol{\varepsilon}}_{i,t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^\tau \varepsilon_{i,t-1-\tau}\right]. \qquad (79)
\end{aligned}
$$

For a SEC process, (85) becomes equal to

$$
E\left[\Delta\widetilde{\mathbf{y}}_{i,t-1} y_{i,t-1}\right] = \frac{1}{1+\lambda}\sigma_f^2\widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'(\boldsymbol{\phi}_i \odot \mathbf{w}_i). \qquad (80)
$$

For an SMA process (85) equals

$$
\begin{aligned}
& E\left[\Delta\widetilde{\mathbf{y}}_{i,t-1} y_{i,t-1}\right] = \widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'E\left[\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\boldsymbol{\varepsilon}_{t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^\tau \boldsymbol{\varepsilon}'_{t-1-\tau}\right](\boldsymbol{\phi}_i \odot \mathbf{w}_i) + \\
& + \widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'E\left[\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\boldsymbol{\varepsilon}_{t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^\tau \varepsilon_{i,t-1-\tau}\right] + E\left[\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\widetilde{\boldsymbol{\varepsilon}}_{i,t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^\tau \boldsymbol{\varepsilon}'_{t-1-\tau}\right](\boldsymbol{\phi}_i \odot \mathbf{w}_i) \\
= \;& \frac{1}{1+\lambda}\sigma_\varepsilon^2\left[\widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'(\boldsymbol{\phi}_i \odot \mathbf{w}_i) + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right]. \qquad (81)
\end{aligned}
$$

In addition, the covariance between $\Delta\widetilde{\mathbf{y}}_{i,t-1}$ and $\Delta y_{i,t-1}$ equals

$$
\begin{aligned}
& E\left[\Delta\widetilde{\mathbf{y}}_{i,t-1} \Delta y_{i,t-1}\right] = E\Bigg[\left(\widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\mathbf{f}_{t-1-\tau} + \sum_{\tau=0}^{\infty}\lambda^\tau \Delta\widetilde{\boldsymbol{\varepsilon}}_{i,t-1-\tau}\right) \\
& \cdot\left((\boldsymbol{\phi}_i \odot \mathbf{w}_i)'\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\mathbf{f}_{t-1-\tau} + \sum_{\tau=0}^{\infty}\lambda^\tau \Delta\varepsilon_{i,t-1-\tau}\right)\Bigg] \\
= \;& \widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'E\left[\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\mathbf{f}_{t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\mathbf{f}'_{t-1-\tau}\right](\boldsymbol{\phi}_i \odot \mathbf{w}_i) + \widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'E\left[\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\mathbf{f}_{t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\varepsilon_{i,t-1-\tau}\right] \\
& + E\left[\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\widetilde{\boldsymbol{\varepsilon}}_{i,t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\mathbf{f}'_{t-1-\tau}\right](\boldsymbol{\phi}_i \odot \mathbf{w}_i) + E\left[\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\widetilde{\boldsymbol{\varepsilon}}_{i,t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^\tau \Delta\varepsilon_{i,t-1-\tau}\right]. \qquad (82)
\end{aligned}
$$

For a SEC process, (82) becomes equal to

$$
E\left[\Delta\widetilde{\mathbf{y}}_{i,t-1} \Delta y_{i,t-1}\right] = \frac{2}{1+\lambda}\sigma_f^2\widetilde{(\boldsymbol{\phi}_i \odot \mathbf{w}_i)}'(\boldsymbol{\phi}_i \odot \mathbf{w}_i). \qquad (83)
$$

For a SMA process, (82) becomes equal to

$$
\begin{aligned}
& E\left[\Delta\widetilde{\mathbf{y}}_{i,t-1}\Delta y_{i,t-1}\right] = \\
= \quad & (\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})' E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\boldsymbol{\varepsilon}_{t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\boldsymbol{\varepsilon}'_{t-1-\tau}\right](\boldsymbol{\phi}_i \odot \mathbf{w}_i) + E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\widetilde{\boldsymbol{\varepsilon}}_{i,t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\boldsymbol{\varepsilon}'_{t-1-\tau}\right](\boldsymbol{\phi}_i \odot \mathbf{w}_i) \\
= \quad & \frac{2}{1+\lambda}\sigma_{\varepsilon}^2\left[(\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})'(\boldsymbol{\phi}_i \odot \mathbf{w}_i) + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right]. \qquad (84)
\end{aligned}
$$

Furthermore, we have

$$
\begin{aligned}
& E\left[\Delta y_{i,t-1} y_{i,t-1}\right] = E\Bigg[\left((\boldsymbol{\phi}_i \odot \mathbf{w}_i)'\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\mathbf{f}_{t-1-\tau} + \sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{i,t-1-\tau}\right) \\
& \cdot\left(\frac{\alpha_i}{1-\lambda} + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)'\sum_{\tau=0}^{\infty}\lambda^{\tau}\mathbf{f}_{t-1-\tau} + \sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{i,t-1-\tau}\right)\Bigg] \\
= \quad & (\boldsymbol{\phi}_i \odot \mathbf{w}_i)' E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\mathbf{f}_{t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\mathbf{f}'_{t-1-\tau}\right](\boldsymbol{\phi}_i \odot \mathbf{w}_i) + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)' E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\mathbf{f}_{t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{i,t-1-\tau}\right] \\
& + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)' E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\mathbf{f}_{t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{i,t-1-\tau}\right] + E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{i,t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{i,t-1-\tau}\right]. \qquad (85)
\end{aligned}
$$

For a SEC process, (85) becomes equal to

$$
E\left[\Delta y_{i,t-1} y_{i,t-1}\right] = \frac{1}{1+\lambda}\sigma_f^2(\boldsymbol{\phi}_i \odot \mathbf{w}_i)'(\boldsymbol{\phi}_i \odot \mathbf{w}_i). \qquad (86)
$$

For an SMA process, (85) becomes equal to

$$
\begin{aligned}
& E\left[\Delta y_{i,t-1} y_{i,t-1}\right] = \\
= \quad & (\boldsymbol{\phi}_i \odot \mathbf{w}_i)' E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\boldsymbol{\varepsilon}_{t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\boldsymbol{\varepsilon}'_{t-1-\tau}\right](\boldsymbol{\phi}_i \odot \mathbf{w}_i) + E\left[\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{i,t-1-\tau}\sum_{\tau=0}^{\infty}\lambda^{\tau}\varepsilon_{i,t-1-\tau}\right] \\
= \quad & \frac{1}{1+\lambda}\left[1 + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)'(\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right]\sigma_{\varepsilon}^2. \qquad (87)
\end{aligned}
$$

Last,

$$
\begin{aligned}
E\left[(\Delta y_{i,t-1})^2\right] &= E\left[\left((\boldsymbol{\phi}_i \odot \mathbf{w}_i)'\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\mathbf{f}_{t-1-\tau} + \sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\varepsilon_{i,t-1-\tau}\right)^2\right] \\
&= \begin{cases} \frac{1}{1+\lambda}\sigma_f^2(\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})'(\boldsymbol{\phi}_i \odot \mathbf{w}_i) + \frac{1}{1-\lambda^2}\sigma_{\varepsilon}^2 & \text{for a SEC process, and} \\ \frac{1}{1+\lambda}\sigma_{\varepsilon}^2\left[1 + (\boldsymbol{\phi}_i \odot \mathbf{w}_i)'(\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right] & \text{for a SMA process.} \end{cases} \qquad (88)
\end{aligned}
$$

Combining (79), (82), (85) and (88) we obtain the following expressions for a SEC and a SMA process, respectively

$$
\begin{aligned}
& cov\left[\Delta\widetilde{\mathbf{y}}_{i,t-1}, (y_{i,t-1} - \widehat{\pi}_2\Delta y_{i,t-1})\right] = cov\left[\Delta\widetilde{\mathbf{y}}_{i,t-1}, \left(y_{i,t-1} - \frac{\sum_i y_{i,t-1}\Delta y_{i,t-1}}{\sum_i (\Delta y_{i,t-1})^2}\Delta y_{i,t-1}\right)\right] \\
= \quad & cov\left(\Delta\widetilde{\mathbf{y}}_{i,t-1}, y_{i,t-1}\right) - \frac{cov\left(\Delta\widetilde{\mathbf{y}}_{i,t-1}, \Delta y_{i,t-1}\right) cov\left(\Delta y_{i,t-1}, y_{i,t-1}\right)}{var\left(\Delta y_{i,t-1}^2\right)} \\
= \quad & \frac{1}{1+\lambda}\sigma_f^2(\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})'(\boldsymbol{\phi}_i \odot \mathbf{w}_i) - \frac{\left[\frac{2}{1+\lambda}\sigma_f^2(\widetilde{\boldsymbol{\phi}_i \odot \mathbf{w}_i})'(\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right]\left[\frac{1}{1+\lambda}\sigma_f^2(\boldsymbol{\phi}_i \odot \mathbf{w}_i)'(\boldsymbol{\phi}_i \odot \mathbf{w}_i)\right]}{\frac{1}{1+\lambda}\sigma_f^2(\widetilde{\boldsymbol{\phi} \odot \mathbf{w}_i})'(\boldsymbol{\phi} \odot \mathbf{w}_i) + \frac{1}{1-\lambda^2}\sigma_{\varepsilon}^2} \neq 0, (89)
\end{aligned}
$$

and

$$
\begin{aligned}
& cov\left[\Delta\widetilde{\mathbf{y}}_{i,t-1},(y_{i,t-1}-\widehat{\pi}_2\Delta y_{i,t-1})\right]= \\
= & -\frac{\left[\frac{2}{1+\lambda}\sigma_\varepsilon^2\left(\widetilde{(\boldsymbol{\phi}_i\odot\mathbf{w}_i)'}(\boldsymbol{\phi}_i\odot\mathbf{w}_i)+(\boldsymbol{\phi}_i\odot\mathbf{w}_i)\right)\right]\left[\frac{1}{1+\lambda}\left[1+(\boldsymbol{\phi}_i\odot\mathbf{w}_i)'(\boldsymbol{\phi}_i\odot\mathbf{w}_i)\right]\sigma_\varepsilon^2\right]}{\frac{1}{1+\lambda}\left[1+(\boldsymbol{\phi}_i\odot\mathbf{w}_i)'(\boldsymbol{\phi}_i\odot\mathbf{w}_i)\right]\sigma_\varepsilon^2} \\
& +\frac{1}{1+\lambda}\sigma_\varepsilon^2\left[\widetilde{(\boldsymbol{\phi}_i\odot\mathbf{w}_i)'}(\boldsymbol{\phi}_i\odot\mathbf{w}_i)+(\boldsymbol{\phi}_i\odot\mathbf{w}_i)\right] \\
\neq & \ 0. \qquad (90)
\end{aligned}
$$

C Proof of Proposition 8

Define $\underline{\boldsymbol{\phi}}_i^o = \phi_i - \boldsymbol{\mu}_\phi$ and $\underline{\alpha}_i^o = \alpha_i - \mu_\alpha = \alpha_i$. Under Assumptions 1-4 and we have $E\left[\underline{\boldsymbol{\phi}}_i^o\underline{\alpha}_i^o\right] = \underline{\boldsymbol{\phi}}_i^o E\left(\underline{\alpha}_i^o\right) = 0$ since ϕ_i is non-stochastic. Furthemore, $Var\left[\underline{\boldsymbol{\phi}}_i^o\underline{\alpha}_i^o\right] = E\left[\underline{\boldsymbol{\phi}}_i^o\underline{\alpha}_i^o\underline{\alpha}_i^o\underline{\boldsymbol{\phi}}_i^{o\prime}\right] = \underline{\boldsymbol{\phi}}_i^o\underline{\boldsymbol{\phi}}_i^{o\prime}\sigma_\alpha^2$ and $Cov\left[\underline{\boldsymbol{\phi}}_i^o\underline{\alpha}_i^o,\underline{\boldsymbol{\phi}}_j^o\underline{\alpha}_j^o\right] = \underline{\boldsymbol{\phi}}_i^o\underline{\boldsymbol{\phi}}_j^{o\prime}E\left[\underline{\alpha}_i^o\underline{\alpha}_j^o\right] = 0$ for $i\neq j$. Hence, from a Weak Law of Large Numbers we have

$$
\frac{1}{N}\sum_{i=1}^N\left[\underline{\boldsymbol{\phi}}_i^o\underline{\alpha}_i^o\right]\xrightarrow{p}0
$$

Furthermore, following an approach similar to (14) we have

$$
\begin{aligned}
& \sqrt{N}\frac{1}{N}\sum_{i=1}^N\left[\underline{\boldsymbol{\phi}}_i\underline{\alpha}_i\right] = \sqrt{N}\frac{1}{N}\sum_{i=1}^N\left[\underline{\boldsymbol{\phi}}_i^o-\left(\overline{\boldsymbol{\phi}}-\boldsymbol{\mu}_\phi\right)\right]\left[\underline{\alpha}_i^o-(\overline{\alpha}-\mu_\alpha)\right]= \\
& \sqrt{N}\frac{1}{N}\sum_{i=1}^N\underline{\boldsymbol{\phi}}_i^o\underline{\alpha}_i^o-(\overline{\alpha}-\mu_\alpha)\sqrt{N}\frac{1}{N}\sum_{i=1}^N\underline{\boldsymbol{\phi}}_i^o-\left(\overline{\boldsymbol{\phi}}-\boldsymbol{\mu}_\phi\right)\sqrt{N}\frac{1}{N}\sum_{i=1}^N\underline{\alpha}_i^o+\sqrt{N}\frac{1}{N}\sum_{i=1}^N\left(\overline{\boldsymbol{\phi}}-\boldsymbol{\mu}_\phi\right)(\overline{\alpha}-\mu_\alpha) \\
= & \ \sqrt{N}\frac{1}{N}\sum_{i=1}^N\underline{\boldsymbol{\phi}}_i^o\underline{\alpha}_i^o+o_p(1) \qquad (91)
\end{aligned}
$$

where the last line follows from the fact that $\left(\overline{\boldsymbol{\phi}}-\boldsymbol{\mu}_\phi\right) = O_p\left(N^{-1/2}\right)$, $(\overline{\alpha}-\mu_\alpha) = O_p\left(N^{-1/2}\right)$, $N^{-1/2}\sum_{i=1}^N\underline{\boldsymbol{\phi}}_i^o = O_p(1)$ and $N^{-1/2}\sum_{i=1}^N\underline{\alpha}_i^o = O_p(1)$. In the same way we have $E\left[\underline{\boldsymbol{\phi}}_i^o\underline{\varepsilon}_{i,t}^o\right] = \underline{\boldsymbol{\phi}}_i^o E\left(\underline{\varepsilon}_{i,t}^o\right) = 0$, $Var\left[\underline{\boldsymbol{\phi}}_i^o\underline{\varepsilon}_{i,t}^o\right] = \underline{\boldsymbol{\phi}}_i^o\underline{\boldsymbol{\phi}}_i^{o\prime}\sigma_\varepsilon^2$ and $Cov\left[\underline{\boldsymbol{\phi}}_i^o\underline{\varepsilon}_{i,t}^o,\underline{\boldsymbol{\phi}}_j^o\underline{\varepsilon}_{j,s}^o\right] = \underline{\boldsymbol{\phi}}_i^o\underline{\boldsymbol{\phi}}_j^{o\prime}E\left[\underline{\varepsilon}_{i,t}^o,\underline{\varepsilon}_{j,s}^o\right] = 0$ for $i\neq j$ and $\forall\ t,\ s$. As a result, $\frac{1}{N}\sum_{i=1}^N\left[\underline{\boldsymbol{\phi}}_i^o\underline{\varepsilon}_{i,t}^o\right]\xrightarrow{p}0$ and $\sqrt{N}\frac{1}{N}\sum_{i=1}^N\left[\underline{\boldsymbol{\phi}}_i\underline{\varepsilon}_{i,t}\right] = \sqrt{N}\frac{1}{N}\sum_{i=1}^N\underline{\boldsymbol{\phi}}_i^o\underline{\varepsilon}_{i,t}^o+o_p(1)$. In addition, $\frac{1}{N}\sum_{i=1}^N\left[\underline{\alpha}_i^o\underline{\varepsilon}_{i,t}^o\right]\xrightarrow{p}0$ and $\sqrt{N}\frac{1}{N}\sum_{i=1}^N\left[\underline{\alpha}_i\underline{\varepsilon}_{i,t}\right] = \sqrt{N}\frac{1}{N}\sum_{i=1}^N\underline{\alpha}_i^o\underline{\varepsilon}_{i,t}^o+o_p(1)$. With these results in mind, the moment conditions given in (43) are equal to

$$
\begin{aligned}
& \text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^N\sum_{t=s+1}^T\underline{y}_{j,t-s}\Delta\underline{v}_{i,t}\right]= \\
= & \sum_{t=s+1}^T\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\left[\left(\frac{\underline{\alpha}_j}{1-\lambda}+\underline{\boldsymbol{\phi}}_j'\sum_{\tau=0}^\infty\lambda^\tau\mathbf{f}_{t-\tau}+\sum_{\tau=0}^\infty\lambda^\tau\underline{\varepsilon}_{j,t-\tau-s}+\theta\sum_{\tau=0}^\infty\lambda^\tau\underline{\varepsilon}_{j',t-\tau-s}\right)\right. \\
& \left.\cdot\left(\underline{\boldsymbol{\phi}}_i'\Delta\mathbf{f}_t+\Delta\underline{\varepsilon}_{i,t}+\theta\Delta\underline{\varepsilon}_{j,t}\right)\right]=0;\ \text{for } s=2,...,T-1, \qquad (92)
\end{aligned}
$$

since $\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\underline{\boldsymbol{\phi}}_i\underline{\boldsymbol{\phi}}_j'\right]=0$ and

$$\begin{aligned}
&\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=s+1}^{T}\underline{y}_{j,t-s}\Delta\underline{y}_{i,t-1}\right]= \\
=\;& \sum_{t=s+1}^{T}\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{\underline{\alpha}_j}{1-\lambda}+\underline{\boldsymbol{\phi}}_j'\sum_{\tau=0}^{\infty}\lambda^{\tau}\mathbf{f}_{t-s-\tau}+\sum_{\tau=0}^{\infty}\lambda^{\tau}\underline{\varepsilon}_{j,t-s-\tau}+\theta\sum_{\tau=0}^{\infty}\lambda^{\tau}\underline{\varepsilon}_{j',t-s-\tau}\right)\right. \\
&\left.\cdot\left(\frac{\underline{\alpha}_i}{1-\lambda}+\underline{\boldsymbol{\phi}}_i'\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\mathbf{f}_{t-1-\tau}+\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\underline{\varepsilon}_{i,t-1-\tau}+\theta\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\underline{\varepsilon}_{j,t-1-\tau}\right)\right]=-\theta\frac{(T-s)}{1+\lambda}\sigma_{\varepsilon}^2. \quad (93)
\end{aligned}$$

For SYS GMM we have

$$\begin{aligned}
&\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta\underline{y}_{j,t-1}\underline{\upsilon}_{i,t}\right]= \\
=\;& \sum_{t=3}^{T}\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\left(\underline{\boldsymbol{\phi}}_j'\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\mathbf{f}_{t-1-\tau}+\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\underline{\varepsilon}_{j,t-1-\tau}+\theta\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\underline{\varepsilon}_{j',t-1-\tau}\right)\right. \\
&\left.\cdot\left(\underline{\alpha}_i+\underline{\boldsymbol{\phi}}_i'\mathbf{f}_t+\underline{\varepsilon}_{i,t}+\theta\underline{\varepsilon}_{j,t}\right)\right]=0, \quad (94)
\end{aligned}$$

and

$$\begin{aligned}
&\text{plim}_{N\to\infty}\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta\underline{y}_{j,t-1}\underline{y}_{i,t-1}\right]= \\
=\;& \sum_{t=3}^{T}\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\left(\underline{\boldsymbol{\phi}}_j'\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\mathbf{f}_{t-1-\tau}+\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\underline{\varepsilon}_{j,t-1-\tau}+\theta\sum_{\tau=0}^{\infty}\lambda^{\tau}\Delta\underline{\varepsilon}_{j',t-1-\tau}\right)\right. \\
&\left.\cdot\left(\underline{\boldsymbol{\phi}}_i'\sum_{\tau=0}^{\infty}\lambda^{\tau}\mathbf{f}_{t-1-\tau}+\sum_{\tau=0}^{\infty}\lambda^{\tau}\underline{\varepsilon}_{i,t-1-\tau}+\theta\sum_{\tau=0}^{\infty}\lambda^{\tau}\underline{\varepsilon}_{j,t-1-\tau}\right)\right]=\theta\frac{(T-2)}{1+\lambda}\sigma_{\varepsilon}^2\neq 0 \quad (95)
\end{aligned}$$

Table 1. Monte Carlo results, $\widehat{\lambda}$ ($RMSE$)

$N = 400$	$\lambda = 0.5$							$\lambda = 0.7$							$\lambda = 0.9$						
	FE	IV	IV*	DIF	DIF*	SYS	SYS*	FE	IV	IV*	DIF	DIF*	SYS	SYS*	FE	IV	IV*	DIF	DIF*	SYS	SYS*
$z = 1/3$, $T = 6$	.161	1.09	.507	.416	.501	.476	.490	.296	.721	.713	.561	.683	.661	.684	.424	7.88	1.01	.560	.788	.844	.876
	(.364)	(25.6)	(.175)	(.226)	(.190)	(.159)	(.109)	(.426)	(1.14)	(.222)	(.309)	(.219)	(.164)	(.112)	(.495)	(307)	(2.92)	(.557)	(.348)	(.165)	(.112)
$z = 1/2$, $T = 6$	.153	.497	.512	.344	.491	.457	.487	.285	.588	.697	.453	.659	.635	.679	.409	1.20	.058	.430	.726	.821	.869
	(.397)	(2.33)	(.300)	(.339)	(.218)	(.219)	(.125)	(.458)	(8.94)	(.971)	(.450)	(.257)	(.223)	(.129)	(.528)	(14.9)	(3.47)	(.693)	(.422)	(.216)	(.127)
$z = 2/3$, $T = 6$	.140	.483	.570	.258	.467	.436	.481	.269	7.43	.739	.342	.616	.611	.669	.389	.134	.736	.351	.649	.797	.858
	(.442)	(3.34)	(2.22)	(.453)	(.251)	(.276)	(.151)	(.504)	(261)	(2.00)	(.574)	(.310)	(.277)	(.155)	(.575)	(21.0)	(7.59)	(.757)	(.487)	(.262)	(.152)
$z = 3/4$, $T = 6$	.132	.565	.574	.212	.404	.423	.475	.258	-3.80	.511	.293	.578	.599	.661	.375	.612	.591	.312	.598	.788	.850
	(.470)	(1.84)	(1.77)	(.507)	(.276)	(.304)	(.171)	(.533)	(215)	(15.3)	(.621)	(.348)	(.303)	(.175)	(.606)	(6.57)	(5.42)	(.785)	(.526)	(.283)	(.171)
$z = 1/3$, $T = 10$	.312	.506	.498	.439	.503	.484	.491	.471	.707	.699	.609	.696	.672	.688	.622	.915	.903	.697	.850	.856	.882
	(.213)	(.207)	(.113)	(.145)	(.107)	(.115)	(.072)	(.247)	(.250)	(.129)	(.176)	(.113)	(.115)	(.072)	(.293)	(2.55)	(.210)	(.320)	(.162)	(.113)	(.070)
$z = 1/2$, $T = 10$	.303	.518	.501	.385	.494	.472	.489	.459	.723	.701	.531	.680	.653	.694	.608	1.23	.913	.591	.814	.835	.878
	(.244)	(.313)	(.160)	(.224)	(.118)	(.165)	(.083)	(.276)	(.663)	(.191)	(.279)	(.130)	(.166)	(.083)	(.318)	(15.2)	(.457)	(.429)	(.207)	(.153)	(.081)
$z = 2/3$, $T = 10$	.291	.533	.534	.320	.477	.457	.485	.443	.718	.705	.447	.652	.634	.678	.589	.570	.935	.515	.759	.816	.870
	(.283)	(.560)	(.524)	(.306)	(.134)	(.212)	(.102)	(.315)	(1.29)	(2.89)	(.374)	(.158)	(.206)	(.102)	(.355)	(16.4)	(13.7)	(.501)	(.267)	(.191)	(.097)
$z = 3/4$, $T = 10$	.283	.832	.497	.287	.464	.449	.481	.433	.773	.560	.409	.631	.625	.673	.576	.970	1.13	.489	.719	.808	.863
	(.307)	(10.4)	(1.58)	(.345)	(.146)	(.235)	(.118)	(.340)	(1.43)	(4.23)	(.414)	(.179)	(.228)	(.118)	(.379)	(13.8)	(15.1)	(.527)	(.305)	(.208)	(.110)

Notes: FE is the fixed effects estimator, IV is the Anderson-Hsiao estimator and DIF and SYS are the first-differenced and system GMM estimators, proposed by Arellano and Bond (1991) and Blundell and Bond (1998) respectively. DIF and SYS are estimated in two steps and they use y_{it-2} and y_{it-3} as instruments for Δy_{it-1} in the first-differenced equations. Furthermore, SYS uses the optimal weighting matrix (when $\sigma_\alpha^2 = 0$), as derived in Windmeijer (2000). The superscript '*' indicates that the corresponding estimator uses instrument(s) with respect to another cross section. The data generating process is given by $y_{it} = \lambda y_{it-1} + \alpha_i + \phi_i f_t + \varepsilon_{it} + \theta \varepsilon_{jt}$, $i = 1, 2, ..., N; t = -48, -47, ..., T$ with $y_{i,-49} = 0$ and the initial 50 observations being discarded. $\alpha_i \sim iidN\left(0, \sigma_\alpha^2\right)$, $\varepsilon_{it} \sim iidN\left(0, \sigma_\varepsilon^2\right)$, $f_t \sim iidN\left(0, \sigma_f^2\right)$ and $\phi_i \sim iidU\left[-0.25, 0.25\right]$. σ_α^2 is chosen to ensure that the impact of the two variance components, α_i and u_{it}, on $var\left(y_{it}\right)$ is held constant. σ_f^2 is normalised to the value of (1) and σ_ε^2 is set according to (66), such that it changes according to the proportion of σ_u^2 attributed to the factor structure. λ alternates between 0.5, 0.7 and 0.9. All experiments are based on $2,000$ replications.

Table 2. Monte Carlo results, $\widehat{\lambda}$ ($RMSE$)

$N = 800$	$\lambda = 0.5$							$\lambda = 0.7$							$\lambda = 0.9$						
	FE	IV	IV*	DIF	DIF*	SYS	SYS*	FE	IV	IV*	DIF	DIF*	SYS	SYS*	FE	IV	IV*	DIF	DIF*	SYS	SYS*
$z = 1/3$, $T = 6$	.162	.534	.501	.427	.508	.478	.496	.295	.750	.703	.569	.701	.663	.693	.424	.968	.885	.548	.867	.846	.888
	(.365)	(.402)	(.120)	(.231)	(.144)	(.159)	(.076)	(.427)	(.795)	(.146)	(.312)	(.164)	(.162)	(.080)	(.495)	(6.07)	(1.94)	(.569)	(.250)	(.163)	(.085)
$z = 1/2$, $T = 6$	.153	.563	.517	.355	.506	.459	.495	.283	.011	.720	.452	.693	.637	.690	.409	.185	.937	.407	.823	.828	.883
	(.401)	(.717)	(.164)	(.353)	(.162)	(.226)	(.088)	(.462)	(21.2)	(.204)	(.474)	(.187)	(.228)	(.093)	(.527)	(37.8)	(.527)	(.717)	(.303)	(.217)	(.098)
$z = 2/3$, $T = 6$	.140	.116	.520	.266	.499	.440	.492	.266	.323	.709	.335	.674	.612	.684	.388	.656	.786	.322	.768	.801	.875
	(.450)	(20.1)	(.498)	(.472)	(.190)	(.291)	(.108)	(.512)	(18.9)	(.709)	(.584)	(.232)	(.292)	(.114)	(.578)	(6.69)	(2.90)	(.788)	(.375)	(.276)	(.120)
$z = 3/4$, $T = 6$	.131	.555	.546	.219	.489	.430	.488	.254	5.75	.771	.281	.653	.600	.679	.374	1.16	.953	.296	.716	.792	.867
	(.481)	(6.64)	(1.55)	(.523)	(.211)	(.323)	(.124)	(.544)	(7.53)	(4.23)	(.638)	(.263)	(.323)	(.131)	(.611)	(13.0)	(4.51)	(.810)	(.430)	(.293)	(.138)
$z = 1/3$, $T = 10$	.313	.505	.501	.442	.505	.486	.499	.468	.741	.706	.614	.700	.674	.697	.618	1.05	.898	.693	.885	.851	.897
	(.216)	(.199)	(.089)	(.148)	(.067)	(.119)	(.054)	(.252)	(.531)	(.094)	(.179)	(.076)	(.119)	(.047)	(.299)	(2.91)	(.137)	(.326)	(.109)	(.117)	(.049)
$z = 1/2$, $T = 10$	.305	.532	.501	.391	.501	.474	.498	.455	.719	.714	.534	.699	.654	.698	.601	1.03	.921	.582	.880	.837	.891
	(.249)	(.472)	(.123)	(.229)	(.078)	(.169)	(.056)	(.279)	(.610)	(.153)	(.281)	(.072)	(.169)	(.057)	(.323)	(12.1)	(.236)	(.436)	(.154)	(.159)	(.059)
$z = 2/3$, $T = 10$	.290	.542	.512	.325	.501	.460	.497	.441	.771	.703	.449	.681	.636	.692	.585	.649	.891	.502	.829	.819	.886
	(.289)	(1.26)	(.211)	(.311)	(.091)	(.219)	(.071)	(.319)	(.994)	(.451)	(.380)	(.101)	(.213)	(.073)	(.358)	(10.5)	(.879)	(.521)	(.211)	(.201)	(.069)
$z = 3/4$, $T = 10$	.280	.741	.525	.290	.493	.451	.492	.430	.815	.731	.404	.672	.629	.689	.572	.961	.939	.473	.765	.810	.878
	(.313)	(3.41)	(.576)	(.365)	(.101)	(.243)	(.081)	(.347)	(2.39)	(1.74)	(.423)	(.124)	(.232)	(.082)	(.384)	(11.2)	(2.22)	(.541)	(.252)	(.219)	(.079)

Notes: FE is the fixed effects estimator, IV is the Anderson-Hsiao estimator and DIF and SYS are the first-differenced and system GMM estimators, proposed by Arellano and Bond (1991) and Blundell and Bond (1998) respectively. DIF and SYS are estimated in two steps and they use y_{it-2} and y_{it-3} as instruments for Δy_{it-1} in the first-differenced equations. Furthermore, SYS uses the optimal weighting matrix (when $\sigma_\alpha^2 = 0$), as derived in Windmeijer (2000). The superscript '*' indicates that the corresponding estimator uses instrument(s) with respect to another cross section. The data generating process is given by $y_{it} = \lambda y_{it-1} + \alpha_i + \phi_i f_t + \varepsilon_{it} + \theta \varepsilon_{jt}$, $i = 1, 2, ..., N; t = -48, -47, ..., T$ with $y_{i,-49} = 0$ and the initial 50 observations being discarded. $\alpha_i \sim iidN\left(0, \sigma_\alpha^2\right)$, $\varepsilon_{it} \sim iidN\left(0, \sigma_\varepsilon^2\right)$, $f_t \sim iidN\left(0, \sigma_f^2\right)$ and $\phi_i \sim iidU\left[-0.25, 0.25\right]$. σ_α^2 is chosen to ensure that the impact of the two variance components, α_i and u_{it}, on $var\left(y_{it}\right)$ is held constant. σ_f^2 is normalised to the value of (1) and σ_ε^2 is set according to (66), such that it changes according to the proportion of σ_u^2 attributed to the factor structure. λ alternates between 0.5, 0.7 and 0.9. All experiments are based on $2,000$ replications.